

# Advancing the theory and applications of Lagrangian Coherent Structures methods for oceanic surface flows

by

Margaux Filippi

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

June 2019

© 2019 Margaux Filippi All rights reserved.

The author hereby grants to MIT and WHOI permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Author .....  
Joint Program in Applied Ocean Science & Engineering, MIT & WHOI  
May 29, 2019

Certified by .....  
Irina I. Rypina  
Associate Scientist in Physical Oceanography, WHOI  
Thesis Supervisor

Certified by .....  
Thomas Peacock  
Associate Professor of Mechanical Engineering, MIT  
Thesis Supervisor

Accepted by .....  
Nicolas Hadjiconstantinou  
Chairman, Committee on Graduate Students, Department of Mechanical  
Engineering, MIT

Accepted by .....  
David Ralston  
Chairman, Joint Committee for Applied Ocean Science & Engineering, MIT





# Advancing the theory and applications of Lagrangian Coherent Structures methods for oceanic surface flows

by

Margaux Filippi

Submitted to the Joint Program in Applied Ocean Science & Engineering, MIT & WHOI  
on May 29, 2019, in partial fulfillment of the requirements for the degree of  
Doctor of Science

## Abstract

Ocean surface transport is at the core of many environmental disasters, including the spread of marine plastic pollution, the Deepwater Horizon oil spill and the Fukushima nuclear contamination. Understanding and predicting flow transport, however, remains a scientific challenge, because it operates on multiple length- and time-scales that are set by the underlying dynamics. Building on the recent emergence of Lagrangian methods, this thesis investigates the present-day abilities to describe and understand the organization of flow transport at the ocean surface, including the abilities to detect the underlying key structures, the regions of stirring and regions of coherence within the flow.

Over the past four years, the field of dynamical system theory has adapted several algorithms from unsupervised machine learning for the detection of Lagrangian Coherent Structures (LCS). The robustness and applicability of these tools is yet to be proven, especially for geophysical flows. An updated, parameter-free spectral clustering approach is developed and a noise-based cluster coherence metric is proposed to evaluate the resulting clusters. The method is tested against benchmarks flows of dynamical system theory: the quasi-periodic Bickley jet, the Duffing oscillator and a modified, asymmetric Duffing oscillator.

The applicability of this newly developed spectral clustering method, along with several common LCS approaches, such as the Finite-Time Lyapunov Exponent, is tested in several field studies. The focus is on the ability to predict these LCS in submesoscale ocean surface flows, given all the uncertainties of the modeled and observed velocity fields, as well as the sparsity of Lagrangian data. This includes the design and execution of field experiments targeting LCS from predictive models and their subsequent Lagrangian analysis. These experiments took place in Scott Reef, an atoll system in Western Australia, and off the coast of Martha's Vineyard, Massachusetts, two case studies with tidally-driven channel flows.

The FTLE and spectral clustering analyses were particularly helpful in describing key transient flow features and how they were impacted by tidal forcing and vertical velocities. This could not have been identified from the Eulerian perspective, showing the utility of the Lagrangian approach in understanding the organization of transport.

Thesis Supervisor: Irina I. Rypina  
Title: Associate Scientist in Physical Oceanography, WHOI

Thesis Supervisor: Thomas Peacock  
Title: Associate Professor of Mechanical Engineering, MIT

# Acknowledgments

This thesis is the outcome of six years in the MIT/WHOI Joint Program. As is typically the case with interdisciplinary studies and field studies in particular, large collaborations are what made the work possible and as such, I would like to thank the people who were involved, as well as the organizations who funded my work and the people who helped me professionally and personally.

My advisor Thomas Peacock first introduced me to the field of dynamical systems theory when I joined his lab in June of 2013. Tom has always been enthusiastic about my work and his excitement was always contagious and motivating. Tom has also trusted me with many responsibilities and enabled many travels around the world; for these incredibly enriching experiences, I am tremendously grateful.

My advisor Irina Rypina officially joined this project in October of 2017. I am eternally grateful for her brilliance, her rigor and her insight, which have shaped a lot of this thesis. Her patience and her commitment to me as a student were also greatly appreciated. I hope the friendship that was developed can endure for many years.

The members of my thesis committee Pierre Lermusiaux and Amala Mahadevan were remarkably involved in my work throughout the years. Their exceptional expertises were humbling and a valuable input to this thesis. They were both also very supportive of me as a mentee and their devotion to teaching has been recognized by many students.

The mentorship of Alireza Hadjighasem made chapter 2 of this thesis possible. I must also thank George Haller for his interest and his useful feedback. The *braids* group, Jean-Luc Thiffeault, Marko Budišić and Michael Allshouse, have also taught me enormously about fluid flows, dynamical system theory and the world of research. Their guidance and their levity have been an incredible help over the past six years. I would also like to thank two other fluid dynamicists who have mentored me professionally and personally: Dick K.P. Yue and John O. Dabiri.

The analyses in Chapter 3 of this thesis were carried out on the numerical model run and output by Matthew Rayson at the University of Western Australia (UWA) in Perth. The field work in this chapter was enabled by Ryan Lowe, Greg Ivey and Carlin Bowyer

at (UWA). The trip was funded by Thomas Peacock and MIT MISTI, as well as by the Martin A. Abkowitz Travel Award from the MIT Mechanical Engineering department. I was funded by the MIT Martin Family Society of Fellows for Sustainability. The cruise was organized by the Australian Institute of Marine Science (AIMS). The principal investigator, James Gilmour, and the whole crew aboard research vessel Solander, were an incredibly help with the drifter release; moreover, their dedication to coral science was inspiring. The team at UWA and at AIMS offered me a warm welcome to Perth and my trip to Western Australia nothing short of amazing. It also included some of the best freediving I have ever experienced, in a remote coral atoll; to all the people who made this possible, I am forever grateful.

Chapter 4 was one of the outcomes of the ALPHA project, an international collaboration funded by the NSF Hazards SEES grant 1520825. Numerous members of the ALPHA team assisted in the releases of drifters and of drogues, so I must extend my thanks to the whole team. Pierre Lermusiaux's MIT MSEAS group provided the numerical model data; Pierre, Patrick Haley and Chinmay Kulkarni in particular were very helpful. Benjamin Hodges at WHOI taught me the building of mock drogues and the preparation of CODE drifters: these lessons will always be remembered. Finally, Siavash Ameli was an incredible help throughout the years: this thesis could not have been completed without his TRACE platform.

The Academic Program Office at WHOI supported me financially for the last year of this thesis. On a personal level, I would also like to acknowledge their encouragement and dedication to students. Thank you to Meg Tivey, Ed Boyle, Leanora Fraser and Henrik Schmidt for the support. In the Applied Ocean Physics & Engineering department, I would like to express my gratitude to Andone Lavery for her mentorship and her support throughout the years. Being part of WHOI was an honor. Thank you also to my fellow JP students and my cohort of Oceans 13.

In the Mechanical Engineering department at MIT, I would like to thank the staff for supporting and cheering on their students. Many thanks in particular to Leslie Regan, Ray Hardin, Lorraine Rabb and Saana McDaniel. To my former and honorary labmates in the ENDLab, including Sasan John Ghaemsaidi, Rohit Supekar, Boyu Fan and Gerald (Jerry) Wang: thank you for the camaraderie. Project Manus and MakerWorkshop, under the

mentorship of Marty Culpepper, have provided me with a home in MechE and made me the engineer I am today. To all my fellow MW mentors, mentees and friends, thank you so much for the teachings and the friendships. You are everything I expected MIT to be. For this life-changing experience, I would like to express my gratitude to my dearest friend Maha Niametullah Haji and to Marcel Thomas.

The pursuit of a doctoral thesis can be grueling and I have been blessed with the support and the love from my friends and family. I would like to express my gratitude to a few people in particular. First, I could never thank my mother enough for everything she has done for me. My father made me who I am and I will always cherish what he has taught me. To my sister, Carole, thank you for looking after me and for comforting me through the ups and downs. To my step-father, Eric, thank you for the supportive and the invariably positive attitude. Maha Niametullah Haji has been here for me at every step of this thesis. To her, and to my friends Lauren Kuntz and Jackie Cristina Diaz-Sua, thank you for carrying me through graduate school. You are the best friends I could ask for. Finally, I am forever grateful to Ernest C. Browne IV for his unconditional love and support. By helping me build drogues, driving me between campuses, flying to conferences and constantly motivating me, you have facilitated so much over the past couple of years. Thank you for making my life so beautiful.



# Contents

<b>1</b>	<b>Introduction</b>	<b>21</b>
1.1	The importance of ocean surface transport at the submesoscale . . . . .	22
1.2	The Lagrangian versus Eulerian perspectives of transport . . . . .	27
1.3	Common LCS detection methods . . . . .	30
1.3.1	Finite-Time Lyapunov Exponent . . . . .	31
1.3.2	Cluster-based methods . . . . .	32
1.3.3	Non-exhaustive review of LCS methods . . . . .	33
1.3.4	Bickley Jet example . . . . .	36
1.4	Previous applications of LCS to ocean surface flows . . . . .	40
1.5	Thesis overview . . . . .	42
<b>2</b>	<b>A parameter-free spectral clustering approach with noise-based cluster coherence metrics</b>	<b>45</b>
2.1	Motivations . . . . .	46
2.2	Clustering methods considered for Lagrangian Coherent Structures detection	49
2.2.1	Fuzzy C-Means (FCM) . . . . .	51
2.2.2	Conventional Spectral clustering . . . . .	56
2.3	Updated spectral clustering approach with noise perturbation metrics . . . .	66
2.3.1	Deviations from Hadjighasem <i>et al.</i> [2016] . . . . .	67
2.3.2	Coherence metrics . . . . .	70

2.3.3	Algorithm Summary . . . . .	73
2.4	Application to benchmark flows . . . . .	75
2.4.1	Bickley Jet . . . . .	75
2.4.2	Duffing oscillator . . . . .	79
2.4.3	Asymmetric Duffing oscillator . . . . .	86
2.5	Conclusions . . . . .	89
<b>3</b>	<b>Uncovering transport in a coral atoll with Lagrangian Coherent Structures</b>	<b>91</b>
3.1	Motivations . . . . .	92
3.2	Scott Reef geography . . . . .	96
3.3	Numerical modeling of Scott Reef . . . . .	98
3.4	Predictive analysis of the 2007 dataset . . . . .	100
3.5	Field experiments . . . . .	103
3.6	Results . . . . .	105
3.6.1	Summary of the key experimental results . . . . .	106
3.6.2	FTLE analysis of the 2016 dataset: neap tide versus spring tide events	112
3.6.3	Spectral clustering analysis of the 2016 dataset: finding the optimal parameters . . . . .	116
3.6.4	Ocean physics and the role of surface convergence . . . . .	119
3.7	Discussion . . . . .	126
3.8	Acknowledgments . . . . .	129
<b>4</b>	<b>Case studies of Lagrangian Coherent Structures around No Man's Land</b>	<b>131</b>
4.1	Overview . . . . .	132
4.1.1	The ALPHA project . . . . .	132
4.1.2	Oceanography and bathymetry of the domain . . . . .	134
4.1.3	Numerical model . . . . .	134
4.1.4	Experimental methods . . . . .	137



4.2	No Man’s Land 2017 experiment . . . . .	137
4.2.1	LCS predictions . . . . .	139
4.2.2	Experimental results . . . . .	142
4.2.3	Comparison of experimental and numerical trajectories . . . . .	144
4.2.4	LCS analysis . . . . .	146
4.2.5	Discussion . . . . .	152
4.3	No Man’s Land 2018 experiment . . . . .	153
4.3.1	LCS predictions . . . . .	154
4.3.2	Experimental results . . . . .	156
4.3.3	Comparison of experimental and numerical trajectories . . . . .	156
4.3.4	Comparison of experimental and numerical trajectories for the 04:00 to 10:00 time window on August 8 . . . . .	160
4.3.5	LCS analysis for the 04:00 to 10:00 time window on August 15 . . . .	161
4.3.6	Discussion . . . . .	166
4.4	Conclusions . . . . .	168
<b>5</b>	<b>Conclusion</b>	<b>173</b>
5.1	Synthesis of results . . . . .	173
5.2	Discussion . . . . .	175
5.3	Upcoming work . . . . .	179
<b>A</b>	<b>Parameter-free spectral clustering protocol with the maximum distance function</b>	<b>181</b>
A.1	Bickley Jet . . . . .	181
A.2	Duffing oscillator . . . . .	183
<b>B</b>	<b>FTLE ridge calculations for Scott Reef</b>	<b>187</b>
<b>C</b>	<b>Details on Lagrangian computations for the Martha’s Vineyard case study</b>	<b>197</b>

C.1	2017 No Man’s Land Experiment . . . . .	197
C.1.1	Finite-Time Lyapunov Exponent (FTLE) . . . . .	197
C.1.2	Spectral clustering . . . . .	198
C.1.3	Encounter volume . . . . .	202
C.2	2018 No Man’s Land Experiment . . . . .	204
C.2.1	Spectral clustering . . . . .	204
C.2.2	Encounter volume . . . . .	205

# List of Figures

1-1	A chart of the Gulf Stream by Poupard & Franklin [1786]. . . . .	22
1-2	Satellites images of ocean eddies and vortices highlighted by plankton blooms.	24
1-3	British Petroleum oil discarded into the Gulf of Mexico from the Deepwater Horizon spill. . . . .	25
1-4	A rotating saddle misclassified as a vortex by most nonobjective diagnostics, from Haller [2015]. . . . .	28
1-5	Steam rings blown by Mount Etna. Photography courtesy of Tom Pfeiffer. Figure from [Haller, 2015]. . . . .	30
1-6	Comparison of Lagrangian methods on the Bickley jet example. Figure adapted from Hadjighasem <i>et al.</i> [2017]. . . . .	37
1-7	Encounter volume for the periodic Bickley jet flow using encounter radius $5 \times 10^5$ . Figure adapted from Rypina & Pratt [2017] . . . . .	37
1-8	Comparison between the evolution of fluid patches, LCS and drifters. Figure from Olascoaga <i>et al.</i> [2013]. . . . .	41
1-9	Comparison between FSLE and drifters. Figure from Haza <i>et al.</i> [2010]. . . .	43
2-1	Examples of LCS methods targeting leakage-free vortices. . . . .	48
2-2	Comparison between K-Means and spectral clustering on six two-dimensional datasets. Figures from the Python <i>scikit</i> documentation [Pedregosa <i>et al.</i> , 2011]. . . . .	50

2-3	Sensitivity of FCM to initial random initialization: output of 5 iterations of the FCM algorithm on the same Bickley jet example dataset. . . . .	55
2-4	Mock example for spectral clustering with nodes $x_i$ , the edges $e_i$ between them, the matrix $W$ of the associated weights $w_{ij}$ and the sparsified $W$ . . . .	58
2-5	Mock example of the partition of a set of nodes from figure 2-4, the graph Laplacian and the corresponding eigenvalues. . . . .	60
2-6	The impact of the value of the radius $r$ on the size of the detected clusters and the number $k$ of connected components with a Bickley Jet example. . . .	62
2-7	The impact of the distance function on spectral clustering with a Bickley Jet example. . . . .	65
2-8	Example spectral clustering results for the Bickley Jet with runs iterated from Hadjighasem <i>et al.</i> [2016]. . . . .	66
2-9	Illustration of the gap ratio. . . . .	69
2-10	Illustration of the spectral clustering protocol with coherence metrics. . . . .	72
2-11	Mean pairwise distances between particles and boundary filtering examples. .	74
2-12	Bickley Jet flow reiterated from Rypina <i>et al.</i> [2007]. . . . .	76
2-13	Quasi-periodic Bickley Jet flow reiterated from Hadjighasem <i>et al.</i> [2016]. . .	77
2-14	Step 1 of the spectral clustering protocol for the Bickley Jet example: gap ratio as a function of $r$ . . . . .	78
2-15	Step 2 of the spectral clustering protocol for the Bickley jet. . . . .	79
2-16	Step 3 of the spectral clustering protocol for the Bickley jet: results with coherence metrics under noise. . . . .	80
2-17	The Duffing oscillator. (a) Poincaré map with 1000 periods of perturbation $T_{pert} = 2\pi/\omega$ . (b) Forward time FTLE for $30T_{pert}$ , time chosen for the spectral clustering analysis. (c) FTLE ridges in forward (red) and backward (green) time for $10T_{pert}$ . (d) Superimposed forward (positive, red tones) and backward (negative, green tones) FTLE fields for $10T_{pert}$ . . . . .	81

2-18	Steps 1 and 2 of the spectral clustering protocol for the Duffing oscillator. . .	83
2-19	Step 3 of the spectral clustering protocol for the Duffing oscillator: results with coherence metrics under noise. The FTLE ridges are superimposed and correspond to the boundaries of the clusters. . . . .	84
2-20	Boundary calculations for the single connected set within the Duffing oscillator, detected with $r = 2.0$ . . . . .	84
2-21	The asymmetric Duffing oscillator. (a) Poincaré map with 1000 periods of perturbation $T_{pert}$ . (b) Forward and (c) backward FTLE for $10T_{pert}$ . . . . .	86
2-22	Steps 1 and 2 of the spectral clustering protocol for the asymmetric Duffing oscillator. . . . .	88
2-23	Step 3 of the spectral clustering protocol for the asymmetric Duffing oscillator: results with coherence metrics under noise perturbation. The FTLE ridges are superimposed and correspond to the boundaries of the clusters. . . . .	89
2-24	Coherence metrics applied to closed orbits of the Poincaré map of the asymmetric Duffing oscillator. . . . .	90
3-1	Connectivity matrix for the Hawaiian archipelago. Retrieved from NOAA Coral Reef Watch based on the works of Kolinski & Cox [2003]; Kool <i>et al.</i> [2010] . . . . .	93
3-2	Connectivity and LCS in Coral Bay, Australia. Figures and captions extracted from Leclair <i>et al.</i> . . . . .	95
3-3	Location of the Scott Reef atoll system off the northwest coast of Western Australia (maps accessed in ESRI, ArcGIS). Figure and caption from Foster & Gilmour [2018]. . . . .	96
3-4	Bathymetry of Scott Reef from Rayson <i>et al.</i> [2018]. . . . .	97
3-5	Meshes used for the SUNTANS model runs. . . . .	100
3-6	FTLE analysis of the 2007 dataset for Scott Reef South. . . . .	101

3-7	(a) Spectral clustering and (b) FTLE analyses of the 2007 dataset for the channel. Data from Alireza Hadjighasem. . . . .	102
3-8	Research Vessel Solander from the Australian Institute of Marine Science docked in Broome, Western Australia. . . . .	104
3-9	Surface drifters built <i>ad hoc</i> . . . . .	104
3-10	Release strategy for the drifters at Scott Reef . . . . .	106
3-11	Scott Reef field release: drifter trajectories. . . . .	107
3-12	Scott Reef field release Part 1. Comparison between the experimental drifter positions, the velocity field and the LCS analysis at the start of the release at (Left) 06:30 and (Right) 07:00. . . . .	110
3-13	Scott Reef field release Part 2. Comparison between the experimental drifter positions, the velocity field and the LCS analysis throughout the release at 07:30, time at which all drifters were in water; 10:00, roughly half-way through the 6-hour experiment; 12:00 at low tide and start of the flood; and 13:30, 6 hours after all drifters were in water. . . . .	111
3-14	FTLE fields in forward and backward times for Scott Reef South around (a) neap tide on September 27 and (b) spring tide on October 3. . . . .	113
3-15	FTLE fields in forward and backward times in the Scott Reef channel around (a) neap tide on September 27 and (b) spring tide on October 3. . . . .	115
3-16	Periodicity of the FTLE ridges around spring tide: superimposition of FTLE ridges (dark red) at 6 consecutive (a) highs and (b) lows of the tidal cycle. .	116
3-17	Step 1 of the spectral clustering protocol for the Scott Reef channel. . . . .	118
3-18	Step 2 of the spectral clustering protocol for the Scott Reef channel. . . . .	119
3-19	Step 3 of the spectral clustering protocol with coherence metrics for the Scott Reef channel. . . . .	120

3-20	Spectral clustering results for the Scott Reef high tide event with a 12-hour integration window, area shrinkage and vertical fluxes for the northern cluster over 12 hours. . . . .	121
3-21	Divergence in the Scott Reef channel before and after removing the divergent component for (a) the neap tide event and (b) the spring tide event. . . . .	123
3-22	Forward FTLE fields in the Scott Reef channel for the non-divergent velocity fields around (a) neap tide on September 27 and (b) spring tide on October 3. . . . .	124
3-23	Step 1 of the spectral clustering protocol for the non-divergent velocity fields: high tide during (a) neap tide and (b) spring tide. . . . .	125
3-24	Step 3 of the spectral clustering protocol for the non-divergent velocity field at high tide around neap tide: results with coherence metrics for the peak at $r = 0.025$ . . . . .	125
4-1	Google Earth satellite image of the region of interest around the island of Martha's Vineyard, Massachusetts. . . . .	133
4-2	Bathymetry south of Martha's Vineyard. Depths in meters. Data from MSEAS. . . . .	135
4-3	CODE/DAVIS-type drifters. . . . .	138
4-4	LCS predictions for the August 14, 2017 No Man's Land experiment. . . . .	140
4-5	CODE drifter trajectories for the August 14, 2017 No Man's Land experiment. . . . .	143
4-6	(a) CODE drifter trajectories (thick blue lines) and numerical trajectories individually seeded for each ensemble of the model (colored dotted lines). The release positions are shown in red crosses. (b) Pairwise distances between the experimental and numerical trajectories, for each ensemble and each drifter. (c) Same as (b) minus drifters 4-6. . . . .	145
4-7	LCS results for the August 14, 2017 No Man's Land experiment. . . . .	147
4-8	LCS analysis for the August 14, 2017 No Man's Land experiment: zoom in on the clustering results. . . . .	151
4-9	LCS predictions for the August 7, 2018 No Man's Land experiment. . . . .	155

4-10	CODE drifter trajectories for the August 7, 2018 No Man’s Land experiment from 16:00 to 22:00. . . . .	157
4-11	(a) CODE drifter trajectories (thick lines) and numerical trajectories (thin lines), individually seeded in TRACE from the MSEAS velocity fields. Each color correspond to a different drifter trajectory. (b) Pairwise distances be- tween the experimental and numerical trajectories for each drifter. The colors match those in (a). The average distance is plotted with the thick grey line. . . . .	158
4-12	(a) Wind data used by MSEAS forecasts. (b) Measured wind time series for August 7, 2018. . . . .	159
4-13	Tidal time series for August 7 and August 8, 2018. . . . .	161
4-14	CODE drifter trajectories (a) until August 15 12:00, 20 hours after the last drifter deployment and (b) from 04:00 to 10:00 on August 15. The drifter positions are color-coded by time. . . . .	162
4-15	(a) CODE drifter trajectories (thick lines) and numerical trajectories (thin lines), individually seeded in TRACE from the MSEAS velocity fields, for 04:00 to 10:00 on August 15. (b) Pairwise distances between the experimental and numerical trajectories for each drifter. The average distance is plotted with the thick grey line. . . . .	163
4-16	LCS analysis for the August 8, 2018 No Man’s Land experiment. . . . .	167
4-17	Zoom on the coherent spectral cluster at 10:00 from figure 4-16. . . . .	168
A-1	Steps 1 and 2 of the spectral clustering protocol for the Bickley jet with the maximum distance function. . . . .	182
A-2	Step 3 of the spectral clustering protocol for the Bickley jet with the maximum distance function. . . . .	183



A-3	Steps 1 and 2 of the spectral clustering protocol for the Duffing oscillator. (Top) Sweep of $r$ parameters with offset coefficient $10^{10}$ and the average distance function. (Middle) Same for the maximum distance function. (Bottom) Sweep of offset coefficients $10^n$ for the gap ratio peaks. . . . .	184
A-4	Spectral clustering results for the Duffing oscillator with the maximum distance function. . . . .	184
B-1	Comparison between backward-time FTLE ridge advected from 07:30 and computed backward-time FTLE fields on a sliding window covering the Scott Reef field experiment. . . . .	189
B-2	Comparison between backward-time FTLE ridge advected from 07:30 and computed backward-time FTLE fields on a sliding window covering the Scott Reef field experiment (Part 2) . . . . .	190
B-3	Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment.	191
B-4	Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 2). . . . .	192
B-5	Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 3). . . . .	193
B-6	Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 4). . . . .	194
C-1	Tidal time series for August 14, 2017. . . . .	198
C-2	Evolution of the FTLE field at different times of the tidal cycle for the day of August 14, 2017. . . . .	199

C-3	Sweep of $r$ radii for the parameter-free spectral clustering protocol used for the analysis of the 2017 No Man's Land experiment. (a) Sweep for the 14:20 to 20:20 time window. (b) Sweep for the 15:51 to 21:51 time window. . . . .	200
C-4	Sweep of offset coefficients for the $r$ values corresponding to the peaks in gap ratios from figure C-3. (a) Sweep for the 14:20 to 20:20 time window. (b) Sweep for the 15:51 to 21:51 time window. . . . .	201
C-5	Encounter volume over the entire domain of computations for the analysis of the 2017 No Man's Land experiment. (a) Results for the 14:20 to 20:20 time window. (b) Results for the 15:51 to 21:51 time window. (c) Results for the 21:51 to 03:51 time window. . . . .	203
C-6	Sweep of $r$ radii for the parameter-free spectral clustering protocol used for the analysis of the 2018 No Man's Land experiment. Sweep for the 04:00 to 10:00 time window. . . . .	204
C-7	Sweep of offset coefficients for the $r$ value corresponding to the peak in gap ratios from figure C-6. Sweep for the 04:00 to 10:00 time window. . . . .	204
C-8	Encounter volume over the entire domain of computations for the analysis of the 2018 No Man's Land experiment. (a) Results for the 04:00 to 10:00 time window. (b) Results for the 10:00 to 16:00 time window. . . . .	205

# Chapter 1

## Introduction

Ocean surface transport has been at the core of societal problems since the American Revolution: Benjamin Franklin, who charted the Gulf Stream as early as 1768, shared knowledge of its circulation patterns with America's French allies, thereby giving them a maritime tactical advantage in the North Atlantic during the Revolutionary War. A 1786 version of this chart provided in figure 1-1 on page 22. Centuries later, transport in the ocean is still a topic of active research as many environmental disasters are related to ocean surface transport, including the spread of marine plastic pollution, the Deepwater Horizon oil spill and the nuclear contamination from the Fukushima Daiichi plant. Understanding and predicting flow transport, however, remains a scientific challenge.

Building on the recent emergence of Lagrangian methods and other analytical tools from dynamical system theory, this thesis investigates the present-day abilities to describe and understand the organization of flow transport at the ocean surface, including the abilities to detect the underlying key structures, the regions of stirring and regions of coherence within the flow. The motivations for studying ocean transport and, particularly, surface transport are presented in section 2.1. Section 1.2 introduces the Lagrangian perspective to describing flow transport. Section 1.3 presents an introduction to Lagrangian Coherent Structures (LCS). Previous applications of LCS to oceanic flows are reviewed in section 1.4. Lastly, the



Figure 1-1: A chart of the Gulf Stream by Poupard & Franklin [1786]. Knowledge of the surface currents helped navigation of the North Atlantic.

overview of this thesis is presented in section 2.2.

## 1.1 The importance of ocean surface transport at the submesoscale

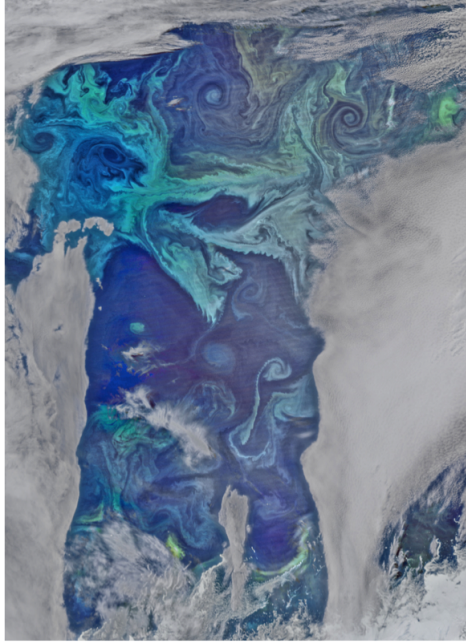
As seawater flows throughout the global ocean, it advects its physical properties and tracers such as heat, nutrients and oxygen. The distribution of different water properties by ocean transport occurs throughout the water column: at the surface, in the deep ocean and at the intermediate upper ocean layer, the layer between the surface and the steep temperature gradient called the thermocline. The dynamics of ocean transport play a major role in regulating Earth's climate [van Sebille *et al.*, 2018]. Ocean physics also impact marine ecology by shaping the environmental conditions around different ecosystems, including their access

to nutrients and light.

The ocean surface is of particular importance because it is the interface between the atmosphere and the ocean: air-sea interactions dictate processes such as heat transport [Thomas *et al.*, 2008; Zhang *et al.*, 2014], which are an essential component of climate models. Moreover, about 50% of oxygen on Earth results from photosynthesis by phytoplankton in the upper layer of the ocean [Field *et al.*, 1998; Marinov *et al.*, 2008; Mahadevan, 2016], typically located within the upper 100 meters. Satellites images of plankton blooms, such as the ones provided in figure 1-2 on page 24, exemplify the role of oceanic structures in transporting and confining nutrients and plankton [Sandulescu *et al.*, 2007]: the chlorophyll marks the edges of the ocean eddies and vortices trapping the plankton.

Ocean surface transport is also at the core of several societal and environmental disasters [Poje *et al.*, 2014]. As previously mentioned, knowledge of the Gulf Stream helped the American Revolution; a 1786 chart was provided in figure 1-1. In 2010, the Deepwater Horizon spill released four million barrels of oil into the Gulf of Mexico, according to an estimate from Crone & Tolstoy [2010], highlighting the importance of reliable forecasts of oceanic contaminant transport [Olascoaga & Haller, 2012]. Pictures of the spill are shown in figure 1-3 on page 25. The following year, an earthquake and a tsunami caused the Fukushima Daiichi nuclear power plant to discharge radioactive particles into the ocean: these tracers have then spread throughout the North Pacific Ocean [Rypina *et al.*, 2013, 2014a], further showing the importance of understanding the pathways and barriers to transport in the ocean.

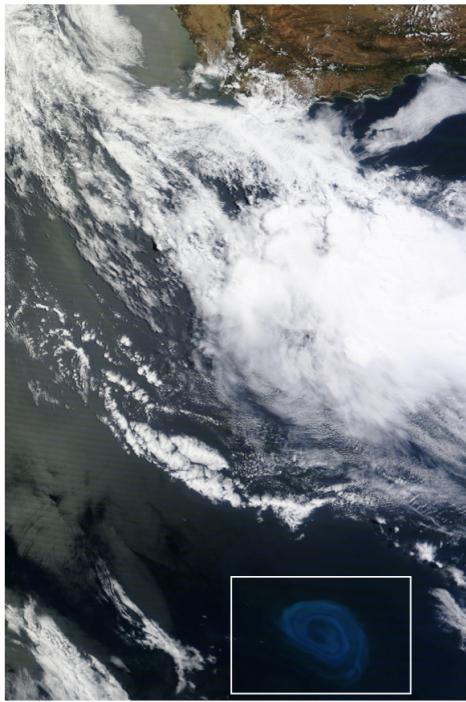
Ocean surface transport is a complex problem because it operates on multiple length- and time-scales that are set by the underlying dynamics. At large length scales, such as that of ocean basins, the ocean dynamics are mostly governed by the Earth’s rotation, by pressure and temperature gradients and by the small depth-to-length ratio, as the horizontal velocities are several orders of magnitude higher than the vertical velocities. At small lengthscales ( $\sim \mathcal{O}(\leq 1\text{m})$ ), the flows are dominated by turbulence and are highly three-dimensional. In



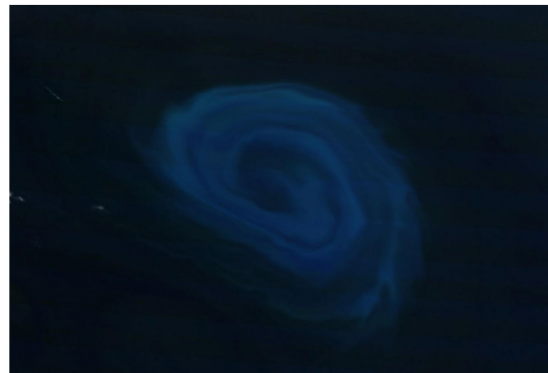
(a)



(b)



(c)



(d)

Figure 1-2: Satellites images of plankton blooms trapped in eddies. (a) Phytoplankton blooms in the Drake Passage stretching about 1300 km and highlighting the multi-scale vortex structures. Image taken with the Visible Infrared Imaging Radiometer Suite on the Suomi NPP satellite (Credits: NASA/Norman Kuring). (b) Phytoplankton blooms (natural-color image) in the Baltic Sea captured by the Operational Land Imager on Landsat 8. (Credits: NASA/Joshua Stevens/U.S. Geological Survey.) (c) 150-km wide eddy off the South Africa coast captured by the Terra satellite with the Moderate Resolution Imaging Spectroradiometer (Credits: NASA/Jesse Allen). (d) Zoom in of the rectangular insert in (c).





(a)



(b)

Figure 1-3: British Petroleum oil discarded into the Gulf of Mexico from the Deepwater Horizon spill. Photographic credits: (a) Beltrá [2010] (b) Kari Goodnough/Bloomberg via Getty Images.

between lie the mesoscale and the submesoscale. Mesoscale motions are characterized by a small value of the Rossby number  $Ro = \frac{U}{Lf}$ , where  $U$  and  $L$  are the characteristic velocity and length scales, respectively,  $f = 2\Omega \sin(\phi)$  is the Coriolis frequency at latitude  $\phi$  and  $\Omega$  is the angular frequency of planetary rotation. Small  $Ro$  means that the flow is in geostrophic balance constrained by the planetary rotation and the pressure gradient force: the Coriolis force dominates over the inertial and centrifugal forces. Typically, the mesoscale refers to scales ranging from 10 to 500 km [Abernathey & Haller, 2018].

At large lengthscales and mesoscales, oceanic flow processes have been extensively studied [Ferrari & Wunsch, 2009; Wunsch & Ferrari, 2018], but within the transition scale between the mesoscale and the small scale, dynamical features are less well understood [Thomas *et al.*, 2008; Poje *et al.*, 2014; McWilliams, 2016]. This transition scale is called the submesoscale: it is often defined dynamically by  $\mathcal{O}(1)$  Rossby number  $Ro$ . Thomas *et al.* [2008] also characterize the submesoscale by  $\mathcal{O}(1)$  Richardson number,  $Ri$ :  $Ri = \frac{N^2}{(du/dz)^2}$ , where  $du/dz$  is the vertical shear in velocity  $u(z)$  and  $N = \sqrt{-\frac{g}{\rho} \frac{d\rho}{dz}}$  is the Brunt–Väisälä buoyancy frequency, with  $g$  the local acceleration due to gravity and  $\rho$  the density. For typical flow speeds of  $0.10 \text{ ms}^{-1}$ ,  $\mathcal{O}(1)$   $Ro$  corresponds to  $\sim 1$  km in Massachusetts and  $\sim 750$  m at very high latitudes;  $Ro$  is undefined at the equator where  $f = 0$ . Typically, the submesoscale corresponds to lengthscales of  $\mathcal{O}(1\text{-}10\text{km})$ .

Submesoscale processes are of particular interest for several reasons: just to name a few, they play a major role in the energy cascade towards smaller-scale dissipation, they are also relevant to primary productivity by phytoplankton [Mahadevan, 2016]. Understanding ocean surface transport at this scale is also of vital importance for the outcomes of search and rescue operations [Peacock & Haller, 2013; Allshouse & Peacock, 2015b], as drifts of  $\mathcal{O}(1\text{-}10\text{km})$  are typically found to occur over 2–3 days. One major reason why dynamical features on the ocean surface are the least understood at the submesoscale is because they present what McWilliams [2016] calls “an observational barrier”: submesoscale processes have shorter spatiotemporal scales than mesoscale or larger features, making it hard to obtain a global



and continuous coverage of these processes with the required spatio-temporal resolution using standard technology and approaches [Zhang & Qiu]. For instance, present-day remote sensing offers global coverage of the Earth, but it only recently started offering the spatial accuracy required to observe submesoscale phenomena globally [McWilliams, 2016]; note that the temporal resolution required for a global coverage of submesoscale processes is still lacking. Smaller features, including fine-scale microstructures of  $\sim \mathcal{O}(\leq 1\text{m})$ , can be measured with shipboard instruments [Seo *et al.*, 2018; Shao *et al.*, 2018; Sun *et al.*, 2018]. Submesoscale processes, however, are too large and too rapidly evolving to be fully captured by typical ship surveys.

This thesis investigates the present-day abilities to describe and understand the organization of flow transport at the ocean surface over the submesoscale range. It is primarily based on a series of field experiments that consisted of drifter releases. The LCS analysis is applied to numerical model data and compared to the experimental investigations, with the goal to evaluate our abilities to detect the underlying key structures, the regions of stirring and regions of coherence within the flow. This analysis is conducted through established LCS methods as well as an approach newly-developed as a part of this thesis, a parameter-free spectral clustering protocol with a noise-based metric to evaluate the coherence of clusters.

## 1.2 The Lagrangian versus Eulerian perspectives of transport

The dynamics of motion in geophysical fluid flows and the patterns of transport can be described from the Eulerian perspective and the Lagrangian perspective. The Eulerian description of fluids represents the flow motion as a function of position  $\mathbf{x}$  and time  $t$  from a fixed reference frame, focusing on the properties of the flow within a domain at an instantaneous time: for example, through the velocity field  $\mathbf{v}(x, t)$ . Velocity is inherently a frame-dependent property and needs to be adapted to the frame of reference to describe the

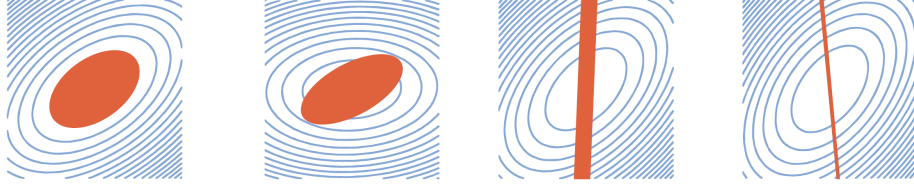


Figure 1-4: A rotating saddle misclassified as a vortex by most nonobjective diagnostics. Figure and caption from Haller [2015].

material behavior. The Lagrangian description, in contrast, represents fluid motion by following fluid parcels as they move through time and space and is a more natural perspective to study ocean transport [Davis, 1983; Mendoza & Mancho, 2012; Lehahn *et al.*, 2018]. Another important consideration to study flow transport is objectivity: the material response should be identified independently of the observer. A nice example that illustrates the need for objectivity is taken from Haller [2015] and provided in figure 1-4 on page 28: the closed and rotating streamlines around the red parcel of fluid can lead to the misclassification of this flow as elliptic or vortical. Advection of the red parcel, however, reveals that it is a rotating saddle flow. Had this parcel of fluid been trapped inside a coherent vortical structure instead, it would have stayed coherent over time.

For a given flow velocity field, the particle trajectories can be generated through the equation of motion:

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(x, t). \quad (1.1)$$

The solutions are denoted by  $\mathbf{x}(x_0, t_0; t)$ , where  $x_0$  is the initial position at initial time  $t_0$  and  $t \in [t_0; t_1]$  is the time instant within the interval from the initial time  $t_0$  to the final time  $t_1$ . This relatively simple equation can nonetheless generate complicated trajectories, even in time-periodic flows: indeed, the resulting trajectories can result in chaotic motion [Aref, 1984; Rypina, 2007]. Chaotic motion is defined by a high sensitivity to initial conditions, meaning that very small changes in initial conditions can result in vastly different trajectories: neighboring particles can separate exponentially with time. Aperiodic flows also commonly generate trajectories with such sensitive dependence on initial conditions [Haller & Poje,

1998]. Oceanic flows are aperiodic, even when they are driven by periodic forcing, such as tidally-driven flows in inlets or coral lagoons. Oceanic velocity fields are typically obtained from analytical models [McWilliams, 1976], numerical models [Deleersnijder & Lermusiaux, 2008; F.J. *et al.*, 2013; Rayson *et al.*, 2018], high-frequency radar systems [Rypina *et al.*, 2014b; Kirincich, 2016], or satellite altimetry [Lehahn *et al.*, 2018] .

The properties of distinct water parcels can be exchanged through homogenization via mixing, following the repeated stretching and folding of material surfaces, such as through mechanical swirling of the fluids. With stirring, the gradients between the water parcels increase sharply. Mixing is the second phase of homogenization and involves molecular diffusion. The processes governing stirring can be interpreted by the field of dynamical systems theory and its advances from the past 15–20 years [Haller & Yuan, 2000; Wiggins, 2005; Villermaux, 2019]. The patterns of transport can be recognized through the detection of areas with high rates versus low rates of stirring or through the detection of barriers to transport that can occur in a flow. Examples of the latter are plentiful in natural flows, including the smoke rings emerging from a puffing volcano, as shown in figure 1-5, oceanic and atmospheric fronts, the eddies shedding from the Gulf Stream or the vortices from figure 1-2. The Lagrangian structures that organize transport and govern coherent trajectory patterns over a given interval of time in such complex fluid flows are referred to as Lagrangian Coherent Structures (LCS), a term coined by Haller & Yuan [2000]. These structures act as the hidden skeleton of spatiotemporally complex fluid flows [Mathur *et al.*, 2007; Peacock & Haller, 2013; Haller, 2015] and are often undetectable and unidentifiable from the direct interpretation of velocity fields in unsteady flows.

This thesis considers several tools used to identify LCS. The following section presents a brief overview of the LCS methods that will be discussed or used in this thesis. Two methods in particular will be further explained in chapter 2: the Fuzzy C-Means (FCM) and the spectral clustering methods, which aim to detect coherent parcels of fluid from clusters of particle trajectories. The other primary LCS tool used is the Finite-Time Lyapunov



Figure 1-5: Steam rings blown by Mount Etna in November 2013. Courtesy of Tom Pfeiffer <http://www.volcanodiscovery.com>. Figure and caption from Haller [2015].

Exponent (FTLE), described in section 1.3.1, whose usability for planning experiments will be assessed in details.

### 1.3 Common LCS detection methods

LCS methods can be classified according to the type of structures they seek to detect. Some methods called cluster-based methods, such as Fuzzy C-Means (FCM) or spectral clustering, seek to detect parcels of fluids that remain coherent and non-filamenting over time, whereas other methods, such as the Finite-Time Lyapunov Exponent (FTLE) or the variational approach, look for the structures delimiting regions with qualitatively different transport behaviors. Many LCS methods require large datasets of Lagrangian trajectories generated by velocity fields. This is the case for the FTLE, the most widely used method that is described in subsection 1.3.1, which requires a greater resolution than, for instance, cluster-based methods. In contrast, some methods, like braid theory, have been developed to specifically analyze sparse datasets of trajectories, as dense datasets of Lagrangian trajectories are sometimes difficult to obtain.

Recent review papers of LCS methods include Allshouse & Peacock [2015b] and Hadjighasem *et al.* [2017]. Two cluster-based methods, FCM and spectral clustering, briefly summarized in subsection 1.3.2, will be further discussed in chapter 2. The encounter vol-

ume method, explained in subsection 1.3.3, will also be used in chapter 4. In chapters 3 and 4 of this thesis, the analysis of the case studies was based on numerical ocean models.

### 1.3.1 Finite-Time Lyapunov Exponent

In 2000, Haller & Yuan coined the term of Lagrangian Coherent Structures (LCS) and defined these structures as the “material lines with locally the longest or shortest stability or instability time”. They computed fields of Finite-Time Lyapunov Exponents, or FTLE, from particle trajectories to look at LCS boundaries. Since then, the FTLE approach has remained the most widely used LCS method. FTLE fields can be computed by the differentiation of the flow map, obtained from numerical trajectories  $\mathbf{x}_j(x_0, t_0; t)$ , typically generated from velocity field datasets with high resolution, over a finite integration time  $t \in [t_0; t_1]$  from initial conditions  $x_0$ . The flow map at time  $t$  is denoted by  $\mathbf{F}_{t_0}^t := \mathbf{x}_j(t_0, x_0; t)$ . The gradient of the flow map is used to compute the Cauchy-Green strain tensor:

$$C_{t_0}^{t_1}(\mathbf{x}_0) = [\nabla \mathbf{F}_{t_0}^{t_1}(\mathbf{x}_0)]^T [\nabla \mathbf{F}_{t_0}^{t_1}(\mathbf{x}_0)] \quad (1.2)$$

as well as its eigenvalues  $\lambda_i(x_0)$ . For a forward-time flow map, the largest eigenvalue yields the largest amount of stretching possible between neighboring tracer particles. It is used to construct the scalar FTLE field and quantify the separation rate within the flow over  $[t_0; t_1]$ .

$$\Lambda_{t_0}^{t_1}(x_0) = \frac{1}{t_1 - t_0} \log \sqrt{\lambda_n(x_0)}. \quad (1.3)$$

Here,  $\Lambda_{t_0}^{t_1}(x_0)$  is the FTLE value at the position  $x_0$  for the integration window  $[t_0; t_1]$ ,  $n$  is the number of dimensions of the trajectories, with  $n = 2$  for a two-dimensional flow, and  $\lambda_n$  is the highest eigenvalue.

Locally maximum values of FTLE that are connected along a curve, referred to as FTLE ridges [Shadden *et al.*, 2005], correspond to structures with the strongest separation rates for forward time advection: the separation of particles for the time window considered is

maximal on each side of the structures. These structures are thus often called ‘repelling’ [Hadjighasem *et al.*, 2017]. Conversely, FTLE ridges of backward-time calculations from  $t_1$  to  $t_0$  are the curves along which the strongest attraction rates occur in forward time and they identify locally attracting features.

For any LCS method, the analysis of a dynamical system depends on the time window of integration from  $t_0$  to  $t_1$ : this time window is a property of the system that is analyzed rather than a parameter of the method. It also depends on the resolution of the numerical grid, i.e., the number of particles included for the computations: it must be sufficient for the method of interest, which can be verified by the convergence of results when the resolution is increased. The FTLE analysis depends on the time window and the numerical grid, but no parameters are needed, making this method very advantageous. It is important to note that the FTLE fields and/or ridges are computed for the chosen window of time  $t_0$  to  $t_1$ . FTLE fields are often computed sequentially over sliding integration windows  $[t_0 + dt; t_1 + dt]$  to look at how the organization of a flow evolves in time. The corresponding FTLE ridges are not time-evolving structures, as they are sequentially recomputed: indeed, each  $[t_0 + dt; t_1 + dt]$  interval represents a different finite-time dynamical system. The sliding-window analysis is however commonly performed [Shadden *et al.*, 2005; Hadjighasem *et al.*, 2017], but it is important to note that the sliding window analysis looks at the LCS within a dynamical system evolving from  $t_0$  to  $t_f$  as opposed to the evolving LCS of the dynamical system at time  $t_0$ . An example of the differences between advected FTLE ridges and sliding-window FTLE computations is provided for the case study in chapter 3, in appendix A-2.

### 1.3.2 Cluster-based methods

Cluster-based LCS methods detect groups of particles that form coherent sets isolated from the rest of the flow. Recently, methods have been developed based on algorithms from unsupervised machine learning, in which all individual particles are assigned to a cluster in an automatic way. The two cluster-based methods, Fuzzy C-Means and spectral clustering,

will be detailed and discussed further in chapter 2 and applied throughout this thesis.

The Fuzzy C-Means (FCM) approach to LCS detection introduced by Froyland & Padberg-Gehle [2015] is a clustering algorithm that assigns particles to clusters based on the Euclidean distance between trajectories. The FCM algorithm allocates the trajectories according to the average over time of the geometrical distance between trajectories and cluster centers. The partition of the domain is optimized when the trajectories are close to their assigned cluster’s center. This optimization is iterated through the particles’ likelihoods of membership to different clusters.

The spectral clustering approach [Hadjighasem *et al.*, 2016] also detects coherent clusters based on the distance between trajectories, but with weights that quantify pairwise similarities between trajectories, using tools from spectral graph theory. Spectral clustering assigns particles to different clusters in order to maximize the intra-cluster similarity while minimizing the inter-cluster similarity. Domains filling the space between coherent clusters correspond to the incoherent cluster, or mixing regions.

### 1.3.3 Non-exhaustive review of LCS methods

One of the first methods developed to detect coherent sets from sparse trajectories was the braid theory approach by Thiffeault [2010]; Allshouse & Thiffeault [2012]. It transposes the time-evolving trajectories into a two-dimensional space-time diagram, which highlights how trajectories intertwine, thereby reducing the data to a sequence of crossings between trajectories. The method measures the rate of topological entanglement of these trajectories and defines coherent groups from sets of particles with minimal entanglement compared to the rest of the dataset. To analyze a set of trajectories over a certain time interval  $T$ , the only parameter to pick is the ratio of entanglement, which distinguishes the coherent sets from the rest of the system where entanglement, and thus mixing, is much higher. However, to yield any result, braid theory necessitates the coherent sets to contain at least a couple of particles with a rate of entanglement much lower than the surrounding region of mixing. The

main drawback of applying the braid theory approach to geophysical flows is that the method requires a certain amount of entanglement of trajectories, as well as a range of entanglement levels that is wide enough to detect coherent group: in open domains, such as oceanic flows, the required coverage of trajectories in space and time is nearly impossible to obtain in field experiments.

The complexity method (CM) was developed by Rypina *et al.* [2011] to sort trajectories within a domain according to their levels of complexities. The method measures the correlation dimension of trajectories and their ergodicity defect  $d$ . The correlation dimension  $c \in [0, 2]$ , where  $c = 0$  corresponds to stationary points,  $c = 1$  to smooth curves and  $c = 2$  to chaotic curves densely covering an area, allows to distinguish relatively complex trajectories from less complex trajectories.  $d$  uses a counting method similar to box counting and measures the deviation from ergodic motion. The trajectory complexity method requires a parameter beyond the grid resolution and the time window of integration: the number of sampled points along the trajectories. It also requires that the boxes in the box counting method decrease from the full domain to a domain small enough.

Rypina & Pratt [2017] also developed the encounter volume method, which computes the volume of fluid that a given trajectory encounters over a finite interval. The encounter volume gives an indication of the mixing potential of the flow: a low encounter volume characterizes, for example, the cores of coherent eddies; a high encounter volume characterizes chaotic regions. The encounter volume is calculated from the number of trajectories passing by a reference trajectory within a threshold distance, the encounter radius, over a time interval  $T$ ; this threshold is the only parameter to be picked by the user for the encounter volume method, along with the numerical grid resolution and the window  $T$ . Another advantage of the encounter volume is that, similarly to the FTLE, it is a visualization of the kinematics of the flow that displays both the regions of low mixing and of high mixing. It reveals where water parcels can exchange water properties and gives an estimate for the mixing potential of the flow. The encounter volume can also be connected to diffusivity [Rypina *et al.*, 2018].



Serra & Haller [2016] also looked at the differentiation of the flow map, but using the rate-of-strain tensor and the initial-time Taylor expansion of the Cauchy-Green strain tensor. In the limit of a small integration window, the Eulerian rate-of-strain tensor governs Lagrangian deformation. The Eulerian Objective Coherent Structures (OECS) are defined as the instantaneous limits of Lagrangian coherent structures. One advantage of OECS is that the computations rely on the velocity field and do not require the generation of trajectories. OECS are frame invariant, but they are not Lagrangian. The only parameter needed is the numerical grid resolution.

The geodesic approach by Haller [2015] applies principles from variational calculus to material surfaces to find extremizing functions. Repelling shrink- or strain-lines and attracting stretchlines are detected from the first and second eigenvectors of the Cauchy-Green strain tensor, respectively, which constitute hyperbolic LCS [Farazmand & Haller, 2012]. Alternating chains of strainlines and stretchlines connecting singularities of the Cauchy-Green tensor form parabolic LCS that minimize Lagrangian shear and are jet cores [Farazmand *et al.*, 2014]. Elliptic LCS are detected through the computation of material-line-averaged stretching; closed material lines for which the stretching is of the same order as neighboring material curves are the elliptic LCS. The outermost closed shear line marks the boundaries of coherent vortices [Haller & Beron-Vera, 2013]. A drawback of the geodesics method is the number of parameters it requires, including, among others, the distance threshold between singularities for computing parabolic LCS.

Rotationally coherent LCS are defined by Farazmand & Haller [2016] as material surfaces whose elements experience identical rotation over a finite time interval. The polar rotation angle (PRA) is computed from the flow gradient  $\nabla F_{t_0}^{t_1}(\mathbf{x}_0)$  to detect the boundaries of rotationally coherent vortices from the outermost closed and convex level curves of the PRA. Building on the rotation angles theory, Farazmand *et al.* [2016] propose to use the Lagrangian-Averaged Vorticity Deviation (LAVD), which is the trajectory-averaged, normed deviation of the vorticity  $\omega(\mathbf{x}(t; x_0))$  from its spatial mean  $\bar{\omega}$ , to identify rotationally coherent

LCS. The choice of the outermost convex contour is somewhat arbitrary.

Similarly to the FTLE, the Finite-Size or Finite-Scale Lyapunov Exponent (FSLE) method [Artale *et al.*, 1997; Aurell *et al.*, 1997] looks at separation in the flow by computing how fast the distance between neighboring trajectories reaches a certain threshold value. The FSLE approach, however, is not as widely used as other LCS methods such as the FTLE: it performs better at larger spatial scales [LaCasce, 2008] and it does not distinguish the different spatial scales of a system. Ultimately, it is unreliably sensitive to the temporal resolution of velocity fields at small scales [Hadjighasem *et al.*, 2017].

### 1.3.4 Bickley Jet example

Most methods mentioned in sections 1.3.1–1.3.3 are illustrated in figures 1-6–1-7 on pages 37–37. The example is the Bickley Jet flow, which will be further detailed in the following chapter. It consists of a zonal jet that is a barrier to meridional transport with three recirculation vortices on each side. The Bickley Jet is a benchmark flow in dynamical system theory because it exhibits transport behaviors that are qualitatively different and are common in realistic ocean flows: the jet, the vortices and the background chaotic zone.

In figure 1-6, panel (a) shows the Poincaré section, which was computed here for the periodic Bickley jet with the parameters as in Rypina *et al.* [2007]. The Poincaré section is a long-established methodology in dynamical system theory to study periodic flows: this stroboscopic mapping plots the trajectory positions at each period. Figure 1-6.(a) reveals how the jet acts as barrier to transport, as well as the presence of vortices corresponding to concentric discretely-sampled closed orbits, which are islands of coherence among the incoherent background. In contrast, the stroboscopic mapping of the particle trajectories outside these closed orbits show clouds of dots corresponding to the chaotic regime of the incoherent background. The six vortices are all of similar sizes.

Panels (b)-(i) in figure 1-6 were taken from Hadjighasem *et al.* [2017], in which the Bickley Jet was modified from Rypina *et al.* [2007] to become quasi-periodic. Because the flow is

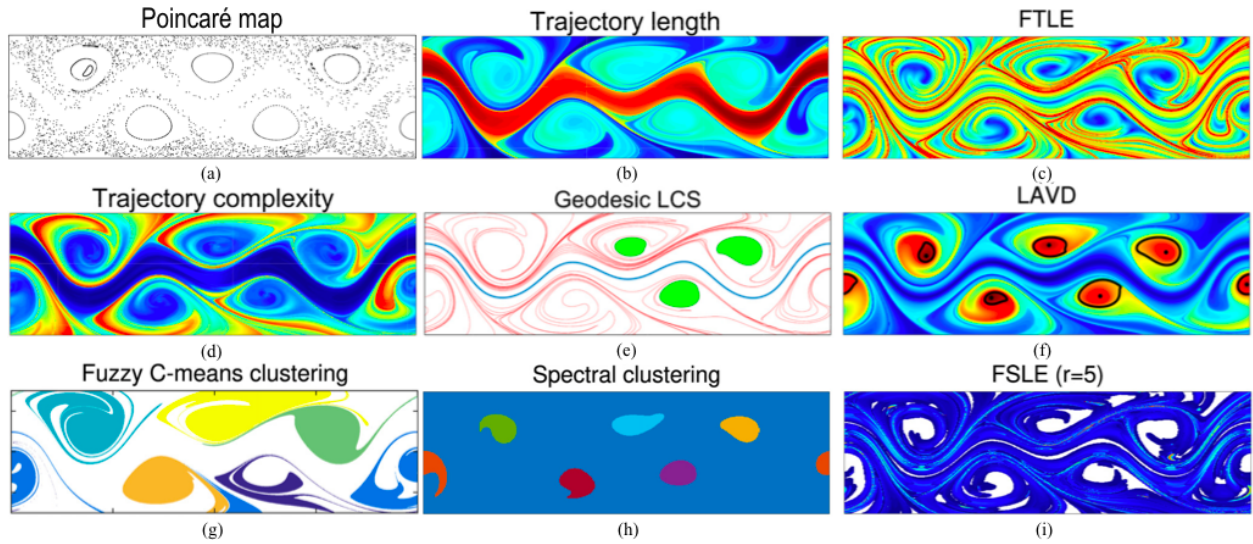


Figure 1-6: Comparison of Lagrangian methods on the Bickley jet example (forward-time calculations only). (a) Poincaré section computed for the parameters in Rypina *et al.* [2007] corresponding to the periodic Bickley Jet flow. (b)-(j) Results of LCS methods on the quasiperiodic Bickley jet example. The panels were adapted from Hadjighasem *et al.* [2017].

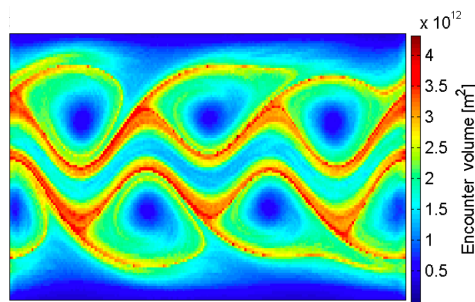


Figure 1-7: Encounter volume for the periodic Bickley jet flow using encounter radius  $5 \times 10^5$ . Figure adapted from Rypina & Pratt [2017]

not periodic, the Poincaré section could not be constructed, but the quasi-periodic system is qualitatively similar to the periodic system. The methods that will be used for analysis in chapters 2–4 of this thesis are the FTLE (figure 1-6.c), FCM (figure 1-6.g), spectral clustering (figure 1-6.h), FCM (figure 1-6.g) and encounter volume (figure 1-7). The flow used to calculate the Poincaré section in figure 1-6.a corresponds to the flow used for the encounter volume method in figure 1-7, the periodic Bickley Jet. The trajectory complexity (figure 1-6.d), the geodesics approach (figure 1-6.e), the LAVD (figure 1-6.f) and the FSLE (figure 1-6.i) are included for completeness and because they are compared with the other methods throughout this thesis, including in section 1.4 and in chapter 2.

Figure 1-6.(b) shows the trajectory length, which corresponds to an arc length computation of the distance covered by the advection particles, as introduced by Mancho *et al.* [2013]. It is not an LCS method *per se*: just as trajectory complexity, FSLE or even FTLE, it does not define what the sought coherent structure are, but it provides an illustration of transport within the domain of interest. Figure 1-6.(b) shows the contrast between the meandering jet, with high values of length in red, and the cores of the recirculation vortices, in light blue. The vortices are hard to discern, however, as there are no clear vortex boundaries.

For the Bickley Jet, various methods captured different aspects of transport within the flow. In the Poincaré section (figure 1-6.a), the closed orbits and the islands of coherence they bound were of similar sizes, suggesting that the vortices detected by LCS method should also be of comparable size. The FTLE scalar field (figure 1-6.c) reveals useful information about the differences in separation rates within the flow, but the individual vortices are hard to discern. The core of the jet shows minimal rates of separation, in blue, consistent with the notion of transport barrier, whereas the boundaries of the jet show maximal rates of separation, in red. The vortices are hardly delimited, however, as the FTLE values smoothly go from blue to green away from the core. While the shape of the jet can be discerned, the vortices are hard to bound. The trajectory complexity (figure 1-6.d) most clearly highlights the meandering jet through its very low values, in dark blue; the vortices are discernible

through relatively low values shown in blue, but the exact boundaries are hard to extract. The geodesic approach (figure 1-6.e) successfully detects the core of the meandering jet as a parabolic LCS, in blue, but only identifies three out of six vortices, the elliptic LCS, in green. The LAVD (figure 1-6.f) method successfully extracts the boundaries of the cores of the vortices, although the cores vary in areas. The jet is also characterized by low values, in blue, but its boundaries are not extractable. The FCM method (figure 1-6.g) can roughly detect the vortices as different clusters, but one vortex is missing its core and the method also includes very distant filaments in the clusters, which contradicts the definition of a coherent structure. These filaments indicate that the coherent sets include trajectories much further than the core of the clusters. The shape and size of the clusters vary highly, in contradiction to the orbits shown in the Poincaré section, which are of similar size. Spectral clustering (figure 1-6.h) successfully detects each of the cores of the recirculation vortices as individual clusters. Moreover, these vortex cores are of similar sizes. The FSLE field (figure 1-6.i) shows the vortices as having low values, but the discrepancy in shape and size of the low FSLE regions make it nearly impossible to extract the coherent structures present in the flow. The jet is hardly discernible. Lastly, the encounter volume method (figure 1-7) also outputs a field that is a representative depiction of the flow transport: the cores of the recirculation vortices have very low encounter volume, in dark blue, consistent with the low mixing potential inside a coherent vortex. The jet core displays a low encounter volume and is delimited by the regions exhibiting high encounter volumes, in red, which is also consistent with high mixing potential.

To summarize, the output of LCS methods generally provides a picture of flow transport within the domain of interest and explains how transport is organized in the flow by revealing its key structures. Different LCS methods emphasize different properties of the flow; for this reason, it is expected that they capture different types of structures. No method is able to detect all the features of interest: here in figure 1-6, no method was able to rigorously extract the jet and the 6 recirculation vortices. Lastly, most methods depend on several user-input

parameters. In chapter 2, this thesis develops an improved parameter-free spectral clustering method. The applicability of different LCS methods, including the chapter 2 parameter-free spectral clustering, to geophysical flows will be assessed in chapters 3 and 4.

## 1.4 Previous applications of LCS to ocean surface flows

LCS are ubiquitous in the ocean: fronts that generate filaments often correspond to hyperbolic structures; elliptic LCS are Lagrangian vortices or eddies; parabolic LCS are shearless cores of jets [Haller, 2015; Hadjighasem *et al.*, 2017]. LCS have been computed in both measured and simulated realistic oceanic flows. They have been shown to underlie the organization of transport of simulated tracers and real drifters.

There are multiple examples in which Lagrangian methods have been used to interpret the observed behavior of surface floaters, or drifters, including: Olascoaga *et al.* [2013]; Jacobs *et al.* [2014]; Beron-Vera *et al.* [2015]; Williams *et al.* [2015]; Rypina *et al.* [2014b, 2016]; Rypina & Pratt [2017]. Typically, in these studies, high numbers of drifters were released and their trajectories were compared to the results from LCS analyses applied a posteriori to velocities from models, satellite altimetry or radar measurements. An example taken from Olascoaga *et al.* [2013] is provided in figure 1-8 on page 41, in which field drifter positions are compared to attracting LCS.

There are very few cases in which Lagrangian methods have actually been used to plan and execute field experiments. Haza *et al.* [2007] and Haza *et al.* [2010] computed FSLE fields from model forecasts for the DART experiment in the Adriatic sea, predicting high relative drifter dispersion. The model was enhanced with high frequency (HF) radar data collected during the drifter experiments. They released surface drifters for a series of consecutive days, with one- or two-day intervals, using the FSLE predictions as a guide. In both studies, the observed trajectories followed a behavior similar to those predicted from the model FSLE fields, as shown in figure 1-9 on page 43. The FSLE, as mentioned in the previous section,

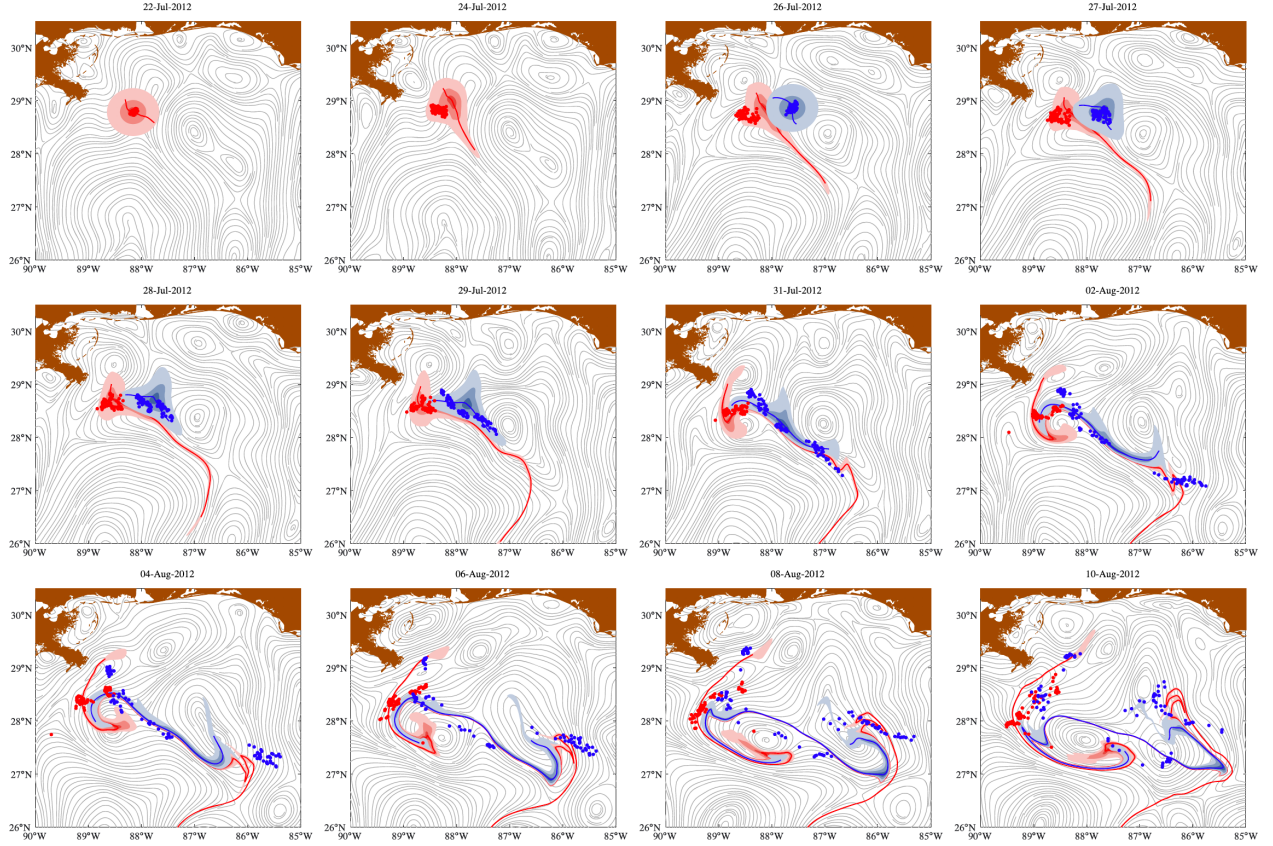


Figure 1-8: Sequence of selected snapshots of the evolution of fluid patches as advected by the altimetry-derived velocity field (4) (red and blue areas) and corresponding centerpieces or attracting LCS obtained as the advected images of forward stretchlines through the patches at the initial time (red and blue curves). Also indicated are S1 (red dots), S2 (blue dots) drifter positions, and instantaneous streamlines of the altimetry-derived velocity field (gray curves). Figure and caption from Olascoaga *et al.* [2013].

is however not widely used for LCS detection anymore as it blends the input timescales [LaCasce, 2008].

To the knowledge of the author, no other field studies have used the by-now standard FTLE method, or other recently-developed set-based methods, to plan and support targeted drifter deployments in regions of distinct transport behavior. The field studies described in the following chapters therefore present the first experiments targeting FTLE ridges and coherent clusters of trajectories, investigating the utility of such Lagrangian methods for operational planning and execution.

## 1.5 Thesis overview

With the recent emergence of a multitude of LCS methods [Allshouse & Peacock, 2015b; Hadjighasem *et al.*, 2017], the primary objective of this thesis is to advance the theory of LCS and then to evaluate the robustness and field applicability of a few common approaches, with a focus on our ability to predict LCS in submesoscale ocean surface flows, given all the uncertainties of the modeled and observed velocity fields, as well as the limitations of the LCS identification methods and the sparsity of Lagrangian data. This includes the design and execution of field experiments targeting LCS from predictive models and their subsequent Lagrangian analysis.

This thesis is organized as follows. A background on LCS and their applications to oceanic flows was given in sections 1.3–1.4. Chapter 2 introduces a new, parameter-free protocol for spectral clustering and proposes a metric for evaluating the coherence of the resulting clusters. The two subsequent chapters are organized around case studies: chapter 3 focuses on a field experiment carried out at Scott Reef, an atoll system in Western Australia, in October 2016, and chapter 4 details two field experiments carried out offshore Martha’s Vineyard, Massachusetts, in August of 2017 and of 2018. For all case studies, the velocity data from numerical models, either forecasting or nowcasting, was provided by collaborators,



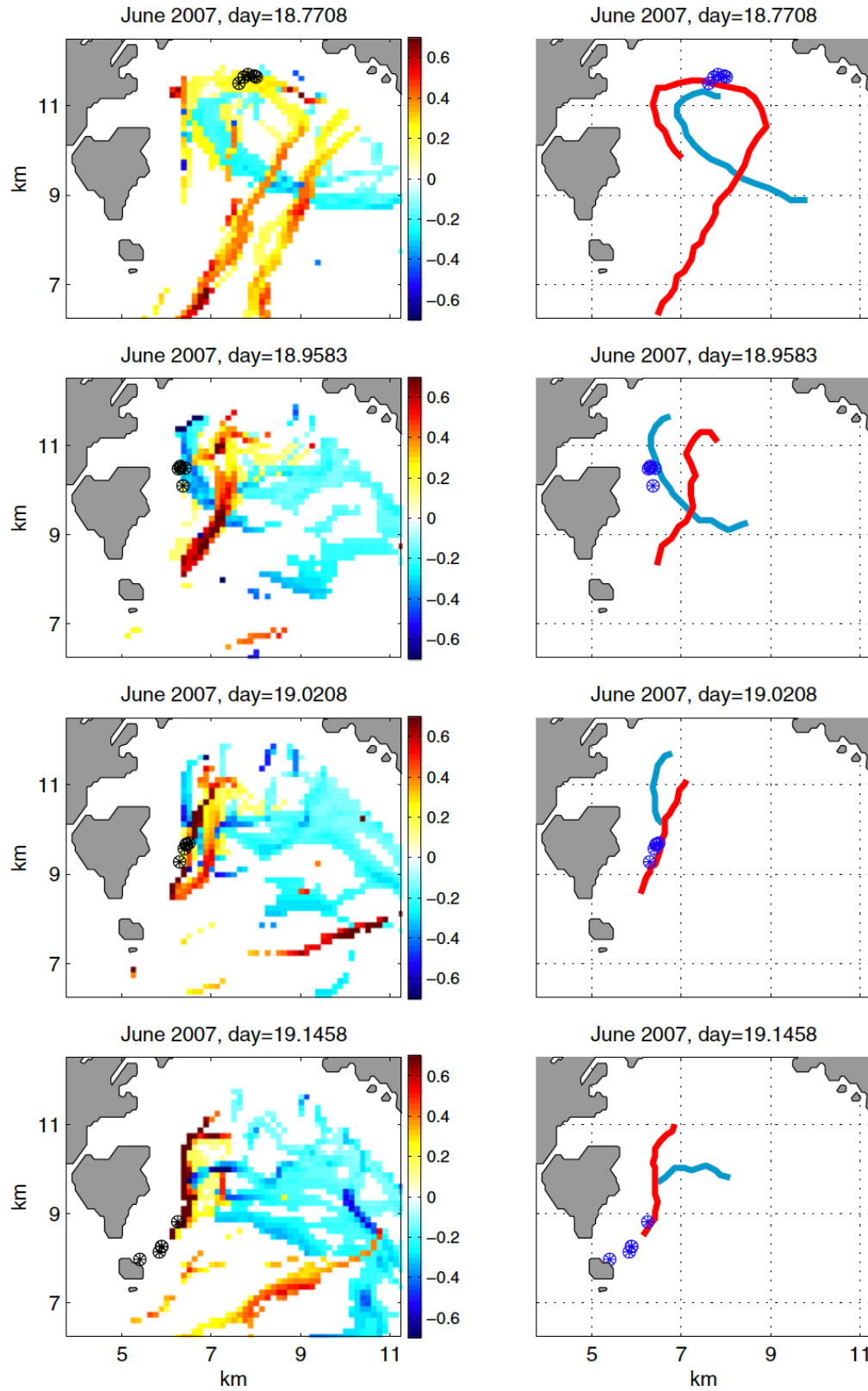


Figure 1-9: (Left panel) Spatial distribution of FSLEs (in per day) computed forward in time (positive) and backward in time (negative) with the observed drifters in cluster 1. (Right panel) The main FSLE ridges that appear to be controlling the drifter trajectories. Figure and caption from Haza *et al.* [2010].

as explained in the respective chapters and in the acknowledgment section. I processed and analyzed the resulting data in the framework of LCS to output predictions of key attracting or repelling LCS structures, as well as the most coherent regions. I planned and executed the bulk of the drifter releases around the predicted structures. In chapters 3 and 3, I will evaluate the comparison between the trajectories and the LCS predictions. Causes for the disagreement will be studied. The methods, models, instrument development and technology used in each case will also be discussed.

## Chapter 2

# A parameter-free spectral clustering approach with noise-based cluster coherence metrics

Recent advances in machine learning have led to an increased popularity of the field, which has been utilized in a very broad range of applications, with other disciplines progressively adopting machine learning methods. The field of Lagrangian Coherent Structures (LCS) has embraced several cluster-based algorithms and adapted the methods to LCS detection. These unsupervised machine learning approaches having emerged so recently, the robustness and applicability of these tools is yet to be proven, especially for geophysical flows. A recurring concern is that the clustering results tend to be highly sensitive to the parameter choices. Building upon the method developed by Hadjighasem *et al.* [2016], an updated spectral clustering approach is presented that avoids arbitrary parameter choice.

The merits and limitations of clustering methods for LCS detection are argued in section 2.1. An overview of commonly used methods is given in section 2.2. Section 2.3 presents the proposed updates to the spectral clustering approach and the noise perturbation metrics to quantify coherence within the clusters. The approach is applied in section 2.4 to three

idealized model flows: the Bickley Jet, the Duffing oscillator and the asymmetric Duffing oscillator. Concluding remarks are presented in section 2.5.

## 2.1 Motivations

A review of commonly-used Lagrangian approaches was given in chapter 1, which presented how LCS methods can be classified into those looking for coherent parcels of fluid and those looking for the structures delimiting regions of different transport behaviors. In their comparison paper, Hadjighasem *et al.* [2017], hereinafter referred to as HA17, classify LCS methods into two groups: diagnostic methods and analytical methods. From the properties of the input dataset of trajectories, diagnostic methods output scalar fields that emphasize the different features of flow transport; analytical methods, on the other hand, use rigorous mathematical principles to define coherence and detect LCS as the solutions of the formulated problems. Analytical methods perform well at detecting the barriers to transport in complex fluid flows: they include the geodesics approach, which has been used for example to detect finite-time mixing barriers in Jupiter’s atmosphere [Hadjighasem & Haller, 2016]; and the Lagrangian-averaged vorticity deviation (LAVD), which has enabled the extraction of eddies in a simulation of the Gulf of Mexico [Beron-Vera *et al.*, 2018]. HA17 include cluster-based methods as analytical methods: these methods assign trajectories to different coherent groups. The clustering algorithms rely on heuristic arguments, however, and so it is considered in this thesis that cluster-based methods are somewhat different from analytical methods, but nevertheless reveal valuable information about the system of interest.

Analytical methods such as the geodesics approach, LAVD or the Polar Rotation Angle (PRA) perform very well for vortices with minimal leakage. Examples of the two methods are provided in figure 2-1 on page 48. The LAVD method in particular was developed to detect the boundaries of Rotationally Coherent Lagrangian Vortices (RCLV). Indeed, the vortex boundaries determined by PRA or LAVD display tangential filamentation only. For

other, non-vortical coherent structures, however, these methods would not necessarily be able to detect structures that stay coherent over  $[t_0; t_1]$  while having their shape deformed or exhibiting some filamentation. LCS that may not exhibit the same behavior as a leakage-free vortex can still be of interest for the study of flow transport and are ubiquitous in the ocean. It is therefore argued in this thesis that cluster-based methods are relevant and useful for the detection and analysis of LCS in geophysical flows, being complementary to more mathematically rigorous methods.

Cluster-based LCS methods, such as Fuzzy C-Means, K-Means or spectral clustering, which are all borrowed from unsupervised machine learning, can provide information about non-vortical yet coherent regions by partitioning the domain into several clusters. A drawback from many machine learning techniques is that the results rely on a set of user-input parameters. The choice of these parameters is often subjective and influenced by the user's biases, which often affects the resulting analysis. HA17 describe how to “choose a reasonable set of parameters for each method” in order to yield “the most favorable outcome”. They describe a robust outcome as an outcome where “small variations in the parameters do not lead to drastic changes in the outcome”, which is a fair way to describe robustness. The reasonability of the parameters and the favorability of the outcome, however, suggest that, in the end, the user will have to make decisions about the outcomes of the analyses that could be arbitrary and subjective. Building on their efforts, this thesis addresses the need for a new protocol to carry on the spectral clustering analysis that, firstly, needs as few user-input parameters as possible and that, secondly, yields unbiased results, in the sense that the protocol should yield optimal results without any need for the user to decide between different outcomes. Optimal results, here, are defined by measuring the coherence of the clusters as robustness under noise; an optimal cluster partition is one where the clusters are the most coherent in comparison to the incoherent background. This metrics for coherence addresses another shortcoming of cluster-based method: once vortices or other isolated islands of coherence occurring in the flow are detected by any of the LCS methods, there is

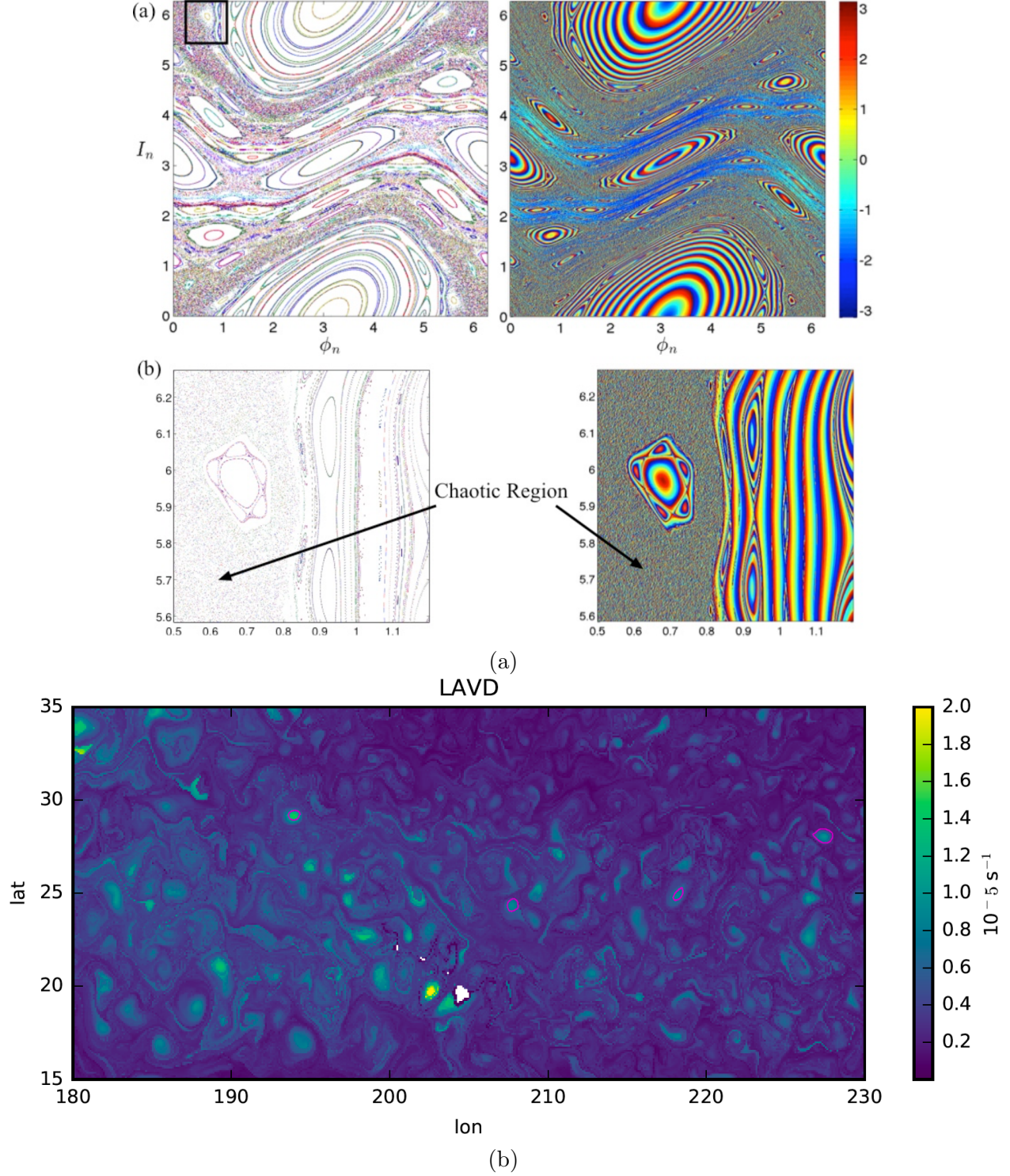


Figure 2-1: Examples of LCS methods targeting leakage-free vortices. (a) Figure from Farazmand & Haller [2016] to illustrate the PRA method. Sub-panel (a) shows the Poincaré map over the torus  $[0, 2\pi] \times [0, 2\pi]$  on the left and the PRA  $\theta_{l_0}^r$  on the right, where polar LCS are closed contours with vortex centers marked by local extrema. Sub-panel (b) shows a close-up view of the region marked by a the rectangle above. (b) Figure from Abernathey & Haller [2018] to illustrate the LAVD method with a close-up example of a 90-day LAVD field from the region near Hawaii. Identified RCLV boundaries are shown as magenta contours.

no established measure of the amount of coherence within these structures. The choice of spectral clustering over other cluster-based methods for our investigations and development will be justified in the following sections.

To recall, the primary goal of this thesis is to evaluate the robustness and applicability of several LCS methods to oceanic surface flows. To this means, the proposed noise-based metric provides a way to evaluate the coherence of the detected sets. The coherence of a set is defined as the robustness of the set to random noise perturbation in order to quantify the coherence of the resulting clusters. This chapter builds upon the spectral clustering approach, previously developed by Hadjighasem *et al.* [2016] for coherent vortex detection specifically, to develop a method to detect coherent sets, not necessarily vortical, in oceanic flows. The emphasis is on ensuring that the modified spectral clustering approach is parameter-free in the sense that the user needs no parameters to use the method.

## 2.2 Clustering methods considered for Lagrangian Coherent Structures detection

The primary cluster-based methods used in the field of LCS are Fuzzy C-Means (FCM) and spectral clustering. Spectral clustering was used to plan the field experiment presented in chapter 3, while FCM was used to plan the field experiments presented in chapter 4. This section presents overviews of the two methods, along with the third related method, K-Means clustering, a special case of FCM, and outlines some pitfalls of the FCM method, which led to the choice of spectral clustering for the proposed parameter-free protocol in section 2.3.

Figure 2-2 on page 50 shows a comparison between spectral clustering and K-Means, a special case of FCM, taken from the Python *scikit* documentation [Pedregosa *et al.*, 2011]. Both algorithms were run on six two-dimensional datasets. K-Means consistently clusters data points according to geometry whereas spectral clustering considers connectivity to cluster the data. This difference is especially obvious in the first example with two concentric

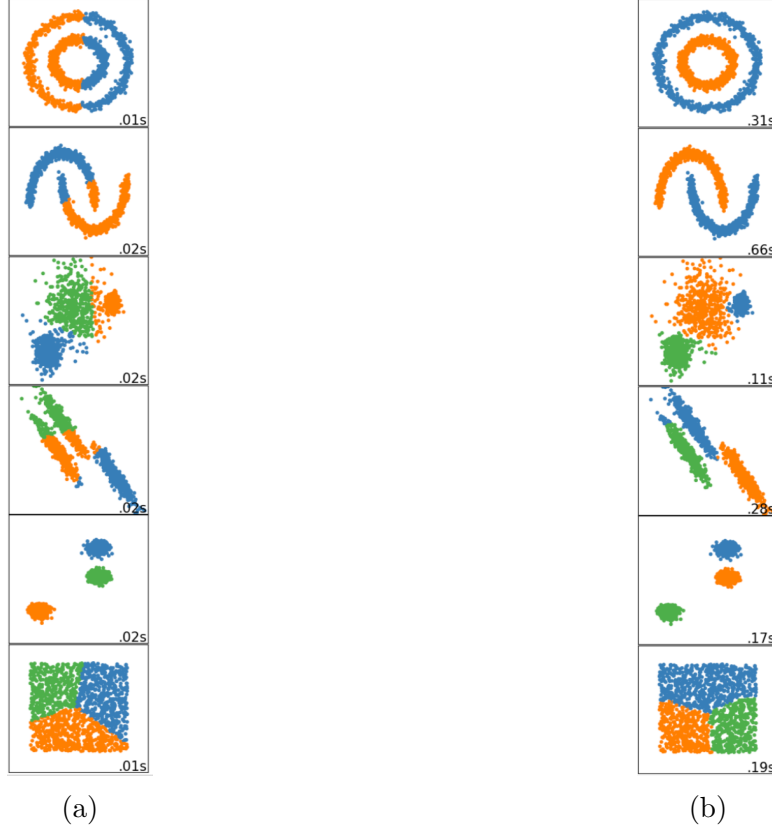


Figure 2-2: Comparison between (a) K-Means and (b) spectral clustering on six two-dimensional datasets. Figures from the Python *scikit* documentation [Pedregosa *et al.*, 2011]. The output clusters are each marked by a different color: orange, blue or green. The times refer to the length of each individual run.

tori, in which K-Means cuts the dataset into a left and a right cluster while spectral clustering picks each torus as an individual cluster. In the second example, each curve is detected as an individual set in figure 2-2.(b), while figure 2-2.(a) has a top/bottom partition. Similarly, in the fourth example, there are three non-continuous patches of points that are each identified as individual spectral cluster (figure 2-2.b), but are divided into top/middle/bottom clusters by K-Means (figure 2-2.a). The third example presents adjacent patches of datapoints: the bigger patch's is divided into a middle and a right cluster by K-Means (figure 2-2.a) but it is detected as an individual set by spectral clustering (figure 2-2.b).



## 2.2.1 Fuzzy C-Means (FCM)

### 2.2.1.1 Overview of the FCM method

For each particle, the FCM algorithm outputs a likelihood, or probability, of membership to each cluster based on the Euclidean distances between trajectories and the cluster centers. For each trajectory, the highest probability of membership will determine which cluster it is assigned to. The method by Froyland & Padberg-Gehle [2015] is followed.

Suppose  $n$  trajectories are given at discrete times, each denoted by:

$$X_i = \{x_{i,0}, \dots, x_{i,T}\} \quad \text{with} \quad x_{i,t} \in \mathbb{R}, \quad 1 \leq i \leq n, i \in \mathbb{N}_*, \quad 0 \leq t \leq T, t \in \mathbb{N}$$

The Euclidean distance  $D(x_{i,0}x_{j,0})$  between two trajectories  $X_i$  and  $X_j$  is based on the  $L_2$  norm and represented by:

$$D(x_{i,0}x_{j,0}) = \sum_{t=0}^T ||x_{i,t} - x_{j,t}||^2 dt = ||X_i - X_j||^2$$

FCM requires the following parameters:

- $K$ , the number of clusters with moving centers:  $C_k = \{c_{k,0}, c_{k,1}, \dots, c_{k,T}\}$ ,  $1 \leq k \leq K, k \in \mathbb{N}_*$
- $u_{k,i}$ , the probability that  $X_i$  is associated with  $C_k$ , where:  $u_{k,i} \in [0, 1] \in \mathbb{R}$
- $m$ , the fuzziness exponent, a relaxation parameter for the shape of the clusters.

The centers are then calculated by:

$$C_k = \frac{\sum_{i=1}^n u_{k,i}^m X_i}{\sum_{i=1}^n u_{k,i}^m} \tag{2.1}$$

And the probabilities of membership by:

$$u_{k,i} = \frac{1 / \|X_i - C_k\|^{\frac{2}{m-1}}}{\sum_{j=1}^K 1 / \|X_i - C_j\|^{\frac{2}{m-1}}} \quad (2.2)$$

The goal is to minimize the ‘fit of membership’ to clusters of the trajectories, measured by the probability-weighted distances within the clusters and represented by:

$$\sum_{k=1}^K \sum_{i=1}^n u_{k,i}^m \|X_i - C_k\|^2 = \sum_{k=1}^K \sum_{i=1}^n u_{k,i}^m \sum_{t=0}^T \|x_{i,t} - c_{k,t}\|^2 \quad (2.3)$$

The algorithm proceeds as follows:

1. Initialize the  $u_{k,i}$ ’s by randomly generating values between  $[0; 1]$ .
2. Calculate the cluster centers  $C_k$ ’s from (2.1).
3. Update the membership probability  $u_{k,i}$ ’s from (2.2).
4. Evaluate the improvement and whether it is below a threshold.
5. Output the clusters with their centers  $C_k$ ’s and the membership values  $u_{k,i}$ ’s for all particles, for all clusters.

This iteration algorithm uses the steepest descent method to converge. The number of dimensions is the number of timestamps per spatial dimensions: if the  $n$  trajectories are in  $R_d$  for  $T$  times, the dimensions of the input matrix are  $T \times d$  and  $T$  centroids are computed for each trajectory.

There are several advantages in utilizing FCM for coherent set analysis. First, it is very fast and easily implemented - there is a built-in MATLAB function *fcm.m*, for instance - so anyone wanting to cluster a dataset can easily carry out the analysis in MATLAB. Second, the membership probabilities indicate how likely particles are to belong to clusters, for each particle and for each cluster, so the ‘fit’ of the different clusters can be compared; for

example, the mean  $u_{k,i}$ 's for each cluster  $k$  can be calculated to look at the different cluster assignments. Third, having these likelihoods of membership, a threshold can be assigned for robustness: for example, we only keep particles with  $u_{k,i} > 75\%$  and discard the particles with lower probabilities to assign them to an incoherent background. Lastly, and perhaps the one main advantage of FCM, each datapoint (i.e. particle or trajectory) is assigned to a cluster.

### 2.2.1.2 K-Means clustering

K-Means clustering is a partitioning method that is very similar to FCM. The main differences are that it partitions the data into  $K$  mutually exclusive clusters and there is no ‘fuzziness’ parameter that come into the calculations of distances – i.e.,  $m = 0$  in equations (2.1–2.3). As each particle is assigned to one cluster only, there is no likelihood of membership output  $u_{k,i}$ , only the *hard* assignment of each particle to its corresponding cluster. For this reason, K-Means is sometimes referred to as ‘Hard C-Means’ in contrast to the *soft* Fuzzy C-means clustering. More in-depth comparisons can be found in Ghosh & Dubey [2013]; Cebeci & Yildiz [2015]. K-Means clustering is of particular interest here because it is the final step in the spectral clustering approach, which will be described in section 2.2.2.

### 2.2.1.3 Challenges of the FCM method

The FCM method contains several shortcomings, many of which are shared with K-Means. First, the number of clusters  $K$  to be input for either algorithm has to be picked by the user. This means the number of sets into which the system is partitioned has to be determined before running the analysis. This parameter is highly subjective, especially in cases where the dynamics of the flow is not known *a priori*, as often happens for geophysical flows. Often times, the analysis is conducted by sweeping through values of  $K$ , running the method several times for each  $K$ , as each run starts with a different random initialization, then by picking the  $K$  value that yields the least amount of variability in the results. This does not always work,

however: consider a domain divided into three neighboring thirds, with the particles located at the boundaries each in the vicinity of at least two clusters. With a random initialization at every run, some of these particles may be assigned to either of their neighboring cluster. This affects the variability of the results, but it does not mean that the cores of the three clusters are not robust or that  $K = 3$  is not the optimal partition.

To illustrate the sensitivity of the FCM method to the random initialization, five iterations of the algorithm were applied to the Bickley Jet using the built-in MATLAB function. The results are shown in figure 2-3 on page 55. The algorithm was run on the same dataset with the exact same parameters and  $K = 7$ , as the flow was known from its streamfunction and its Poincaré section: what was expected was six recirculation vortices within two chaotic zones separated by a meandering jet. The meandering jet is sometimes detected as a coherent structure, but rarely with clustering algorithms, so the total number of expected clusters corresponds to the 6 vortices and the incoherent background as a seventh cluster. In figure 2-3.(a), each cluster is marked by a different color. Figure 2-3.(b) shows the probabilities  $u_{k,i}$  of membership to clusters. In all cases, the vortex at the periodic boundary is split and assigned to two different clusters and the jet is split longitudinally into two or three distinct clusters (turquoise, blue and sometimes yellow): this shows a partition of the flow from geometry only, rather than connectivity or coherence within sets. The vortices are assigned to clusters in groups of two or three, which varied between the runs, proving that the trajectory assignment to clusters highly depends on the initialization. The probabilities  $u_{k,i}$  in figure 2-3.(b) range widely even for vortices within the same cluster. In all cases, the top right vortex has a strong probability of over 95% of belonging to the green cluster, but the bottom right has a much lower probability of belonging to the same green cluster: as low as 50% in the fifth run. For other vortices, such as the top middle vortex, the probability varies highly from run to run and the vortex is assigned to three different clusters out of five runs. For comparison, in the next section, the spectral clustering analysis is run on the exact same dataset as the one in figure 2-3.

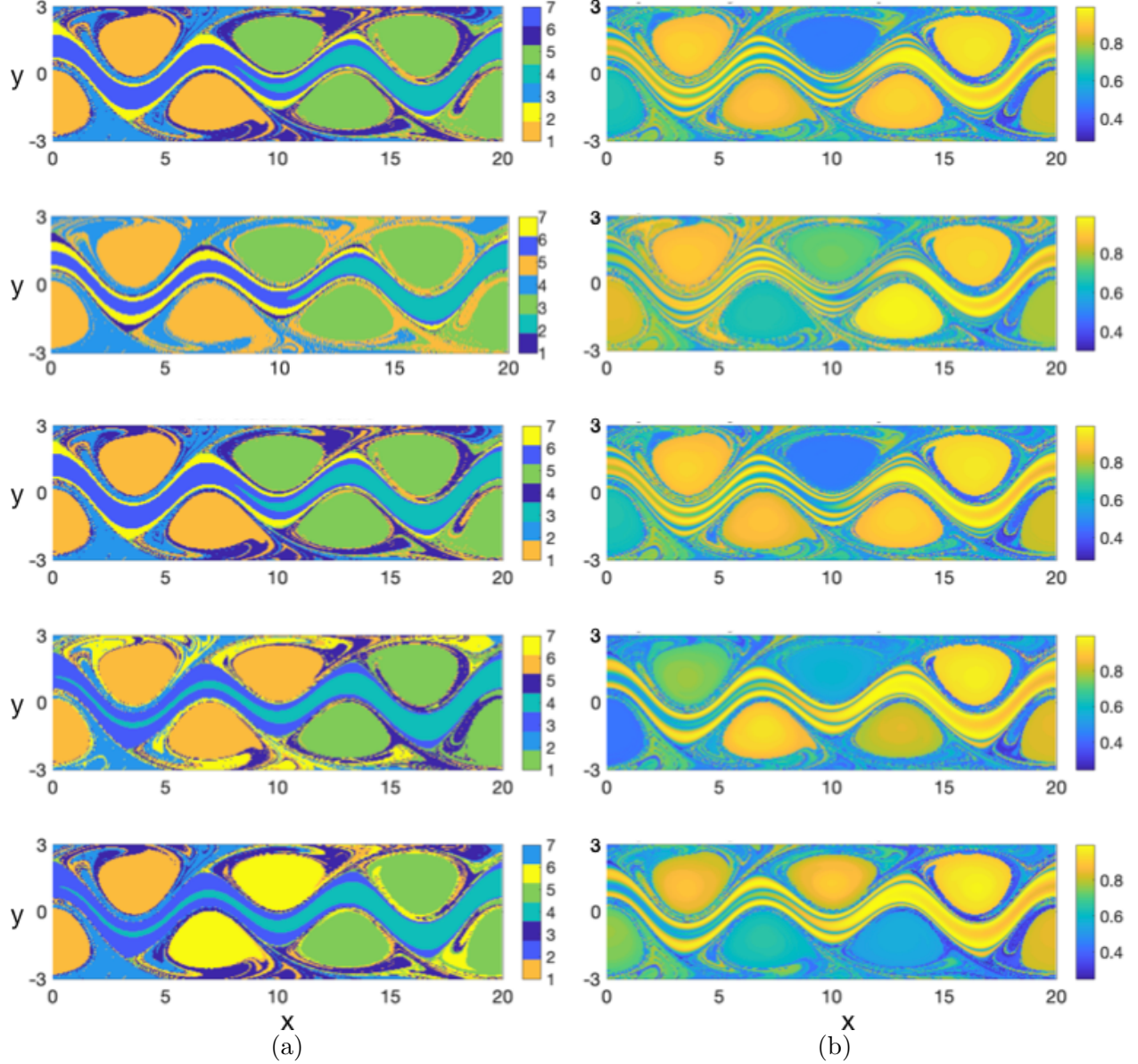


Figure 2-3: Sensitivity of FCM to initial random initialization: output of 5 iterations of the FCM algorithm on the same Bickley jet example dataset. (a) Clusters output, where each cluster is marked by a different color. In all 5 cases, the vortex at the periodic boundary is split and assigned to two different clusters. Runs 1 and 3 are very similar, with clusters containing two (orange) and three (green) vortices. The second run includes some of the incoherent background into the orange cluster. The fourth run outputs a different cluster assignment with the orange cluster contained three full vortices and the yellow cluster being split throughout the incoherent background. In the fifth run, the orange cluster contains one full vortex, the yellow cluster now contains two vortices and the green cluster, two vortices. In all cases, the jet being split into two or three clusters (turquoise, blue and sometimes yellow). (b) Probabilities of membership to clusters. In all cases, the top right vortex has a strong probability of belonging to the green cluster, but for other vortices, the probability varies highly, such as for example for the top center vortex.

A second shortcoming of the FCM and K-Means methods that is related to the arbitrary choice of the number  $K$  of clusters is that the results are highly sensitive to the initialization, which is done by assigning initial probabilities randomly. Vieira & Allshouse [2019] circumvent this issue by computing cluster centers from an educated guess and inputting this value in the first calculation of probabilities. While this reduces the sensitivity to random initialization, it requires an *a priori* understanding of the flow to make this guess. Furthermore, it may introduce subconscious biases.

A third issue with the FCM and K-Means methods is that the algorithms treat each time-step of the trajectory dataset sequentially instead of treating the dataset as such: they require the input data matrix to be in dimensions much larger than the dimensions of the dataset. For example, 100 time-stamps for  $N$  longitude and latitude coordinates correspond to  $N$  2-D trajectories, but for FCM or K-Means, it would correspond to  $200N$  centroids for the initial positions. This problem is commonly referred to as the ‘curse of dimensionality’ in machine learning and is explained in great detail in Aggarwal *et al.* [2001]; Domingos [2012]. This high dimensionality causes non-intuitive results when computing the Euclidean distances at higher dimensions, where the Euclidean distances vary very highly with the position in space and the density of observation points drastically decrease, which reduces the ratio of distances between particles far-away and distances between particles. Moreover, these centroids are artificial and not representative of the dimensions of the trajectories. Constructing a centroid at each step underutilizes the dynamics, as centroids are also artificially advected by the flow.

### 2.2.2 Conventional Spectral clustering

The spectral clustering approach to LCS detection by Hadjighasem *et al.* [2016], hereinafter referred to as HA16, offers a solution to the aforementioned issues for the FCM and K-Means approaches: first, the number of clusters needed for the optimal partition of the data is identified through spectral decomposition; second, the ‘curse of dimensionality’ is avoided and there are no artificial cluster centroids introduced. Spectral clustering partitions

the dataset according to similarity between trajectories such that intra-cluster similarity is maximized and inter-cluster similarity is minimized: the elements within a cluster should be alike, while being different from the elements of other clusters or the elements of the incoherent background cluster. This similarity between elements is determined from the pairwise distances between trajectories. Excellent reviews of spectral clustering and the *Ncut* partition, described below, can be found in Shi & Malik [2000]; von Luxburg [2007].

### 2.2.2.1 Overview of the method

Let  $x^i$  and  $x^j$  be two trajectories in  $\mathbb{R}^d$ , where  $d = [2, 3]$  for two- or three-dimensional flows, respectively, with positions at  $n$  discrete times from  $t_0 < t_1 < \dots < t_{n-1} = T$ . Each position in space and time is denoted by  $x_k^i = x^i(t_k)$ . The distance  $r_{ij}$  between them is:

$$r_{ij} = \frac{1}{t_{n-1} - t_0} \sum_{k=0}^{n-2} \frac{t_{k+1} - t_k}{2} \left( |x_{k+1}^i - x_{k+1}^j| + |x_k^i - x_k^j| \right) \quad (2.4)$$

where  $|\cdot|$  is the spatial Euclidean norm. Next, convert the pairwise distances  $r_{ij}$  into the weights  $w_{ij}$ :

$$w_{ij} = \begin{cases} 1/r_{ij} & \forall i \neq j, \\ \text{constant offset} & i = j. \end{cases} \quad (2.5)$$

From the weights  $w_{ij}$ , build the similarity graph  $G = (V, E, W)$ , with:

- $V = \{x_1, \dots, x_n\}$  the set of nodes, or particles,  $x_i$ ,
- $E \in V \times V$  the set of edges  $e_{ij}$  between nodes  $x_i$  and  $x_j$ ,
- $W \in R^{n \times n}$  the similarity matrix that associates the weights  $w_{ij}$  to the edges  $e_{ij}$ .

An illustrative example is provided in figure 2-4 on page 58.

Following the construction of the matrix  $W$ , the next step is to sparsify  $W$  by removing all weights  $w_{ij}$  below a certain threshold: for a critical distance  $r$ , only keep  $r_{ij} : r_{ij} \geq r$ . This threshold on the distance can be applied over the average distance or it can be more

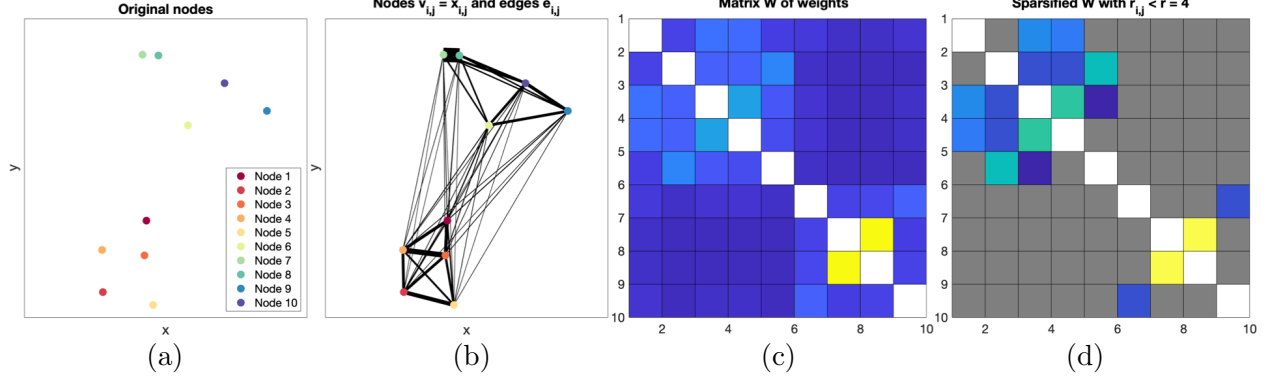


Figure 2-4: Mock example for spectral clustering. (a) A dataset of nodes (particles)  $x_i$  (b) The edges  $e_i$  between the nodes are marked in black and the width corresponds to the associated weights  $w_{ij}$ , which are the inverse distances between  $x_i$ . (c) The matrix  $W$  with entries  $w_{ij}$ . (d) The sparsified  $W$ , where only distances below a threshold  $r$  are kept, i.e., only the weights above a threshold  $1/r$  are kept.

restrictive: the distance between trajectories  $x^i$  and  $x^j$  must be less than  $r$  at all times  $t \in [t_0; t_{n-1}]$ . The advantage of sparsifying  $W$  is two-fold: first and foremost, it saves computational efforts by removing a large amount of entries; second,  $W$  only retains the entries corresponding to a similarity above a critical level, which avoids the case where the dataset partition is influenced by a large number of entries with insignificant similarities. In the figure 2-4 example, for node 1, only the weights  $w_{1,3}$  and  $w_{1,4}$  are kept in the sparsified matrix.

To partition the similarity graph into subsets that maximize intra-cluster similarity and minimize inter-cluster similarity, the partition should be such that the edges within a same group have high weights, meaning the points within a subset are similar to each other, while the edges between different groups have very low weights, meaning the points in different subsets are dissimilar from each other. Note that spectral clustering does not necessarily target the structures with the highest possible coherence value; instead, it partitions the domain such that the coherent structures are most distinct from the incoherent background. This will be demonstrated further in the following section. The incoherent background, or mixing region, or noise cluster, fills the domain between the coherent structures.

The aforementioned goal of partitioning the domain into  $k$  subsets through maximizing



the within-cluster similarity and minimizing the between-cluster similarity is achieved by HA16 using the approximate solutions of the normalized cut function  $Ncut$ . Before defining  $Ncut$ , first define *graph cuts* and *weight cuts*: graph cuts are partitions of the set of nodes  $V$  into subsets. Weight cuts are assigned to graph cuts to evaluate the similarity between the subsets: weight cuts are the sum of the edges weights between sets. For example, for two subsets  $A_1$  and  $A_2$ , the weight cut  $W(A_1, A_2)$  is:

$$W(A_1, A_2) := \sum_{i \in A_1, j \in A_2} w_{i,j} \quad (2.6)$$

To achieve the sought partition of the domain, the graph cut needs to isolate subsets of nodes with high within-group similarity from the rest of the graph: that is, it needs to minimize the weight cuts  $W(A_i, \bar{A}_i)$  between different subsets  $A_i$ 's and their respective complements  $\bar{A}_i$ 's. For  $k$  given subsets of  $V$ , this becomes the minimization problem *mincut* that minimizes the following graph cut:

$$\text{cut}(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k W(A_i, \bar{A}_i) \quad (2.7)$$

In the example provided in figure 2-4, the most logical partition of the graph is to group the first 5 nodes (red-orange-yellow tones) together into a subset  $A$  and the last 5 nodes (green-blue-purple tones) together into a subset  $B$ , as shown in 2-5 on page 60: the  $w_{ij}$  connecting the  $e_{ij}$ 's in  $A$  to the  $e_{ij}$ 's in  $B$  are minimum. In this example,  $\bar{A}$ , the complement of  $A$ , happens to be  $B$ .

This minimization problem is hard to implement in practice because it often separates a few individual nodes from the rest of the graph, but the normalized cut function  $Ncut$  offers a nice solution as it penalizes the smallness of sets. Overall, the  $Ncut$  is an efficient way to partition the set into groups with similarity within but dissimilarity between the different groups Shi & Malik [2000]. The  $Ncut$  can now be defined as follows:

$$Ncut(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{\text{cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)} \quad (2.8)$$

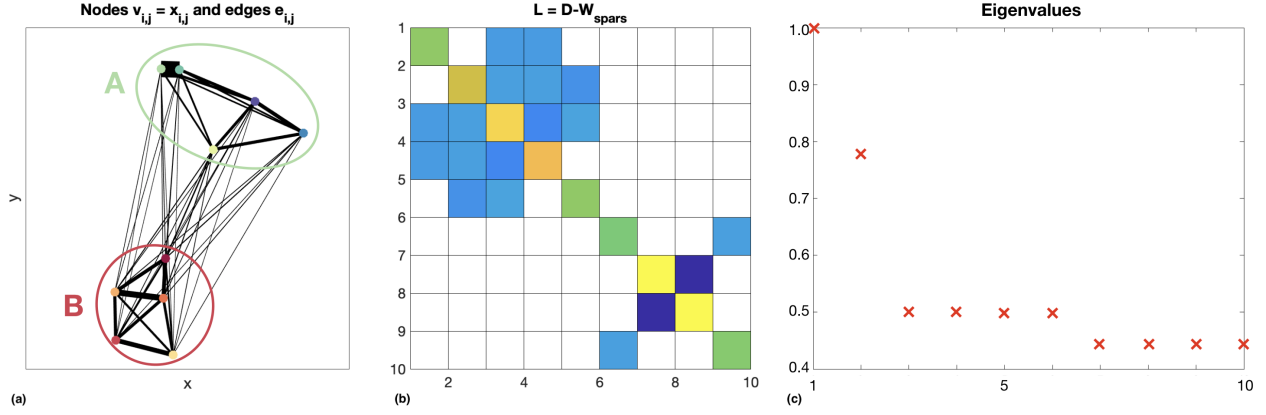


Figure 2-5: Mock example of the partition of a set of nodes (particles) from figure 2-4. (a) The nodes are assigned to two subsets A (green) and B (red). (b) The graph Laplacian  $L$  for the sparsified matrix  $W$ . (c) The plot of eigenvalues of  $L$ , which shows a spectral gap between the second and the third eigenvalues..

where:

$$\begin{aligned} \text{vol}(A) &= \sum_{i \in A} \deg(v_i) \quad \text{is the volume of set } A \\ \deg(v_i) &= \sum_{j=1}^n w_{ij} \quad \text{is the degree of node } v_i \in V \end{aligned} \tag{2.9}$$

For realistic datasets, the *Ncut* problem becomes NP-hard, meaning it is of high computational complexity. It has been shown [Shi & Malik, 2000] that the solution can be approximated by the solution of an eigenvalue problem with the graph Laplacian  $L$ :

$$L = D - W; \quad L\mathbf{v} = \lambda D\mathbf{v}$$

where  $D$  is the diagonal degree matrix built for all  $v_i \in V$  with entries  $\deg(v_i)$ . There can exist  $k$  connected subsets of the graph that are distinctly disconnected from each other, but the between-subset similarity is not exactly zero: hence,  $L$  is a perturbed version of the  $L$  of an ideal case, in which the dissimilarity between subsets is absolute, or the similarity is zero. From perturbation theory and the Davis-Kahan theorem [von Luxburg, 2007], which bounds the differences between eigenspaces of symmetric matrices under perturbations, the

eigenvectors  $\mathbf{v}_i$ 's are close to the vectors from the ideal case. By property of the Laplacian and by invoking the Davis-Kahan theorem, the order of eigenvectors and eigenvalues is meaningful and the first  $k$  eigenvectors  $\mathbf{v}_i$ 's can be used as approximate solutions for the  $k$  connected components of the graph.

Two important considerations are now to, first, retrieve this number  $k$ , and, second, to assess how good this approximation is. Both depend on the spectral gap: the spectral gap is the largest gap between successive eigenvalues of  $L$  and separates the  $k$  connected components from the rest of  $V$ , which is the incoherent subset. This largest gap is from now on referred to as the eigengap. The wider the eigengap, the closer the eigenvectors of the perturbed case are to the eigenvectors of the ideal case: thus, the 'better' the partition of the domain and the more accurate spectral clustering is. The first  $k$  connected components are retrieved by the argument of the eigengap heuristic.

HA16 mention the sensitivity of the results to the critical radius  $r$ , the threshold on distances for the sparsification of the matrix  $W$ : indeed, the spectral decomposition and the size of the eigengap directly depends on  $r$ . This  $r$  value also dictates the number  $k$  of connected components that are detected. The spectral decomposition's dependence on  $r$  is illustrated in figure 2-6 on page 62 for the Bickley Jet example from the previous chapter. The analysis was conducted on a sweep of  $r$  values from 0.5 to 5.0, keeping the rest of the parameters the same as for HA16. Up to  $r = 3.0$ , 6 coherent clusters and 1 incoherent cluster are detected. With  $r = 4.0$ , 7 coherent groups were detected: the 6 vortices plus the meandering jet, bringing the number of clusters to 8. For  $r > 4$ , the number  $k$  decreases: the jet remains detected but the vortices are grouped into pairs, such as the dark blue pair. For  $r = 5.0$ , the yellow cluster englobes the incoherent background and one of the vortices.

Retrieving the value of  $r$  and the set  $U = (v_1, \dots, v_k)$  of the first  $k$  eigenvectors  $v_i$  means that the graph is comprised of  $k$  connected subsets and of the rest of the domain, which is considered to be the incoherent background of particles, meaning that there are ultimately  $K = k + 1$  sub-domains within the graph. Note that for the original set of  $N$  trajectories,  $U \in$

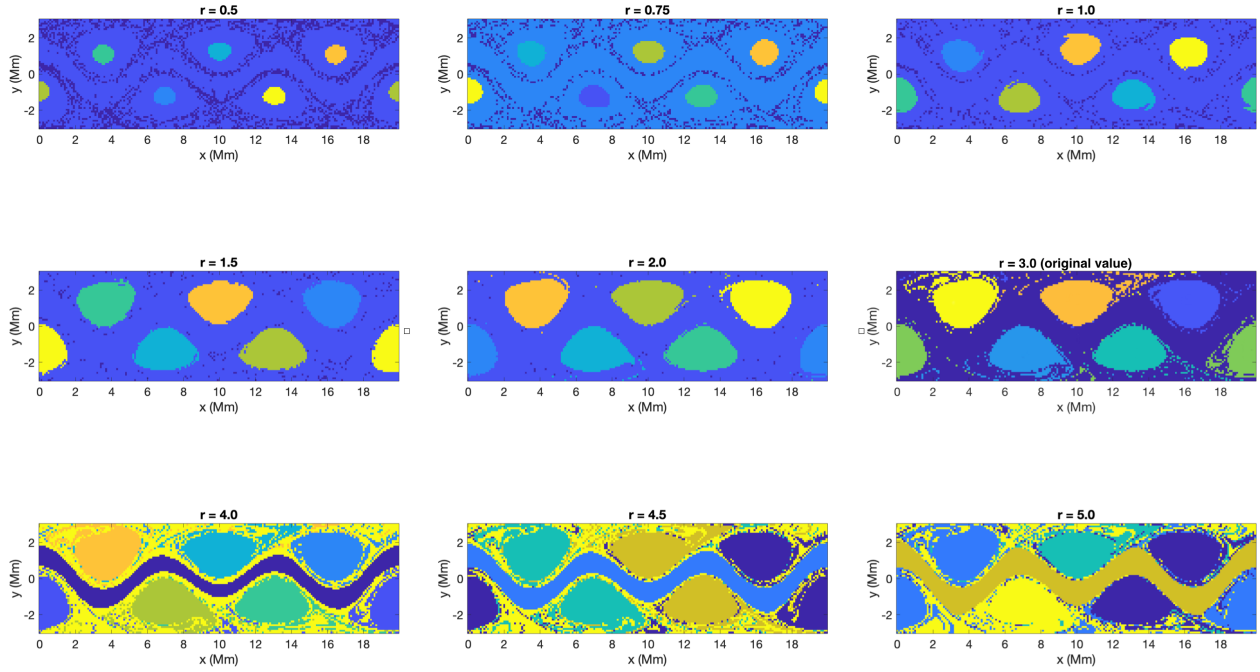


Figure 2-6: The impact of  $r$  on spectral clustering with a Bickley Jet example. Each color corresponds to a different cluster. The size and the number of the coherent clusters are dictated by  $r$ . The value chosen for HA16 was  $r = 3.0$ . With  $r = 4.0$ , 7 coherent clusters were detected: the 6 vortices plus the meandering jet, bringing the number of clusters to 8. For  $r > 4$  values, the jet is detected but the vortices are grouped into pairs. For  $r = 5.0$ , the yellow cluster englobes the incoherent background and one of the vortices.

$\mathbb{R}^{N \times k}$ . Finally, to extract the coherent structures within the domain, the K-Means algorithm is applied to the set  $U$  of connected components using  $K$ : the output is an assignment of the  $N$  trajectories into  $K$  clusters, with  $k$  coherent structures as individual clusters and an incoherent cluster corresponding to the mixing region. K-Means is used because the assignment of the trajectories to their respective cluster is ‘definite’: the graph partition is done such that all particles  $x_i$ ’s assigned to into cluster  $k_n$  have very high similarity, so assigning likelihoods of membership through FCM does not add valuable information. Since the number  $K$  of clusters to retrieve is given by the spectral decomposition, this parameter is thus removed from the list of user-set inputs.

### 2.2.2.2 Challenges of conventional spectral clustering

A clear advantage of spectral clustering over FCM or K-Means for LCS detection is that it yields the number  $K$  of clusters to be detected through the plot of eigenvalues and the space – the eigengap – between them. There are, however, other shortcomings of the method that remain, including user-input parameters that may affect the spectral clustering results. First, the offset in the diagonal of  $W$ , a numerical parameter to avoid  $\rightarrow \infty$  in the diagonal, as seen in equation (2.5), is a scalar value that must be chosen by the user. A second parameter that needs to be input by the user to perform the spectral clustering analysis is the choice of the distance function as a threshold on  $r$ , which significantly impacts the spectral decomposition. For a critical  $r$ , the average distance function puts the threshold on the averages of the pairwise distances between trajectories, as described in equation (2.4). The maximum distance function is more restrictive and in this case, pairwise distances must be less than the critical  $r$  value at all times. This stricter case typically eliminates outside filamentation from the detected clusters.

A comparison of the results from the two functions is provided in figure 2-7 on 65, in which for a specific  $r$  value applied to the Bickley jet, the distance function affects the shape of the clusters and whether the meandering jet is detected. Figure 2-7 displays the plots

of eigenvalues corresponding to each of the clustering results. For both distance functions, the first six eigenvalues are similar, but the difference arises with the seventh eigenvalue and its location in relationship to the maximum eigengap. With the average distance function in panel (a), the seventh eigenvalue appears before the eigengap, whereas with the maximum distance function in panel (b), the seventh eigenvalue is located after. Intuition and inspection of the connected components of the graph reveal that this seventh eigenvalue corresponds to the meandering jet, which appears in the clustering results in (a) but not in (b). Beware, however, that although the correspondence between one of the connected components and the resulting clusters is obvious in this case, there are many counter-examples and, generally, no one-to-one correspondence can be inferred. Note, also, that the order output by the final K-Means assignment is random and the corresponding eigenvalues cannot be retrieved. The implications can be bothersome for several reasons: first, the method outputs clusters, but there is no indication of how coherent they are, as the connectivity cannot be retrieved from the graph. Secondly, the incoherent cluster does not always appear clearly; indeed, for flows that lack strongly connected components, the background cluster is sometimes hard to identify.

Finally, the eigengap heuristic rule to choose the  $r$  value that maximizes the gap between eigenvalues is not always observed. From the regime exploration illustrated in figure 2-8 on page 66, done with the average distance function, the value of  $r = 1.5$  in panel (c) yielded a higher eigengap than the value of  $r = 3.0$  chosen in HA16, shown in panel (b). The offset value, here, was chosen to be  $10^8$ , same as for figure 2-7 but different than for figure 2-6. Note that with  $r = 1.5$  in figure 2-8.c, the vortex boundaries are a bit smaller as the outside filaments are not detected as being part of the structures, which seems to indicate a higher intrinsic coherence.

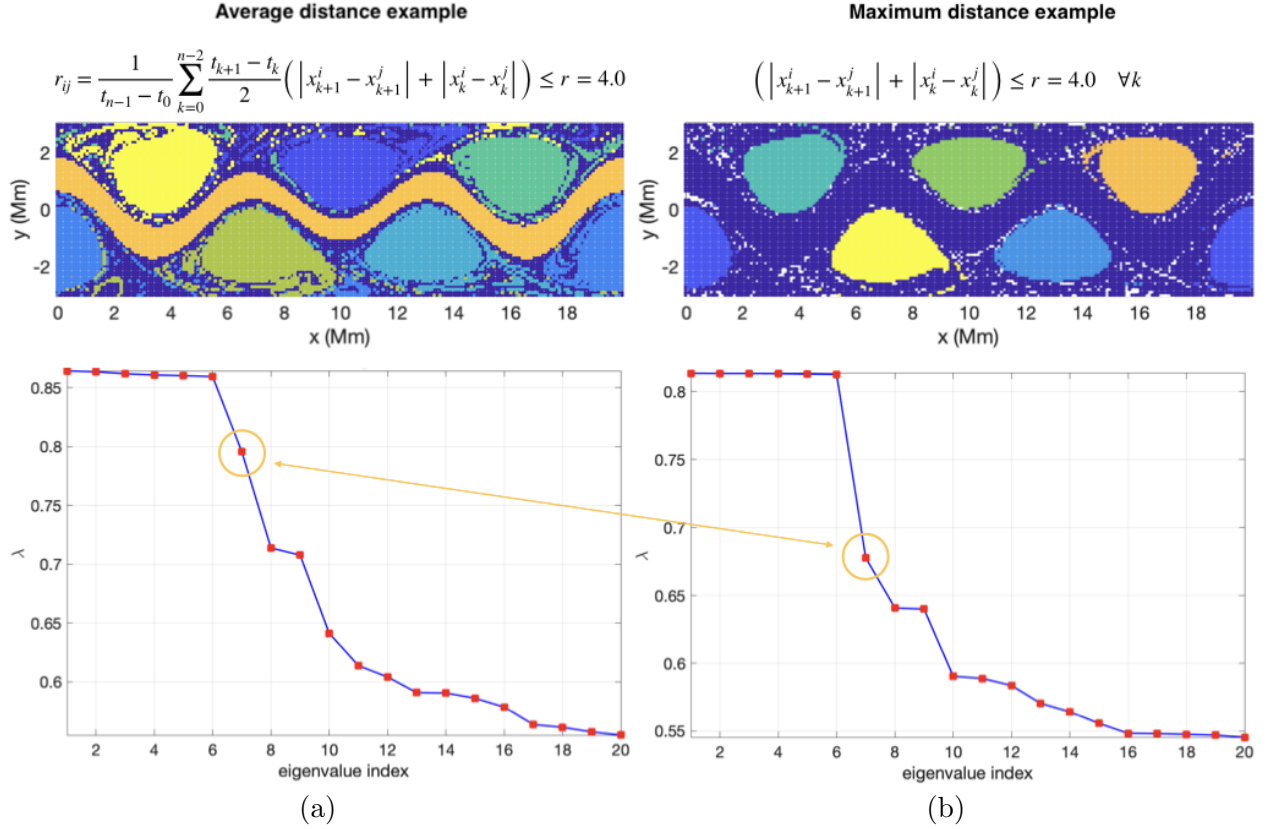


Figure 2-7: The impact of the distance function on spectral clustering with a Bickley Jet example: (a) average distance and (b) maximum distance. Here, the radius was  $r = 4.0$  and the offset value in the diagonal of  $W$  was  $10^8$ . In (a), the meandering jet (orange) was detected as a seventh connected component of the graph. In (b), the more restrictive case on the distance, the  $r_{ij}$  are too large for the jet to be considered a connected component. The filaments outside the vortices are also excluded.

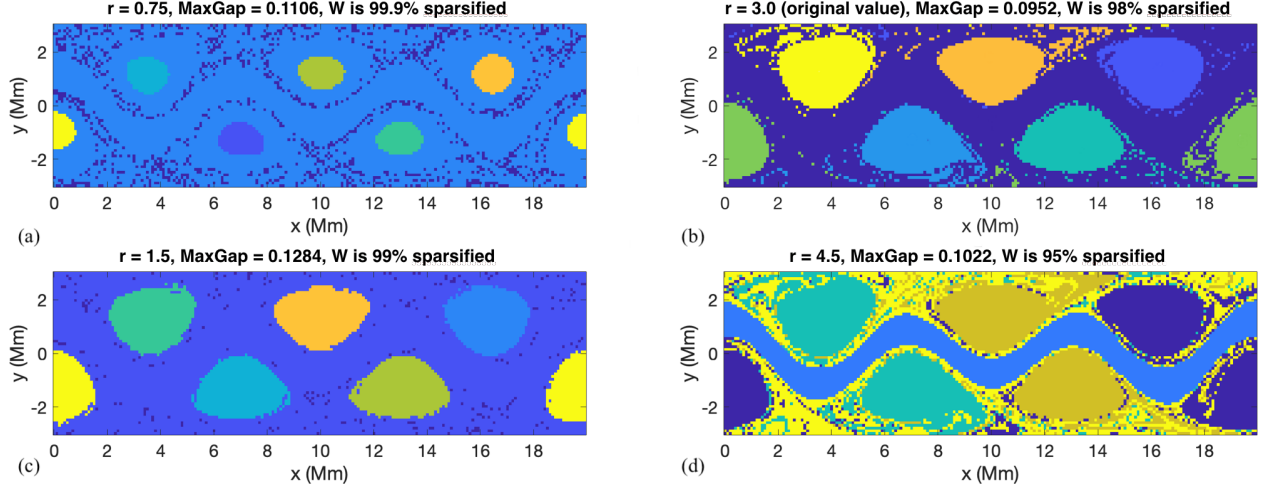


Figure 2-8: Example spectral clustering results for the Bickley Jet with runs iterated from HA16. The originally chosen value was  $r = 3.0$  in panel (b). Three other values of  $r$  yielded a higher eigengap, however. Moreover, to follow the rule of thumb of 5–10% sparsification,  $r = 4.5$  in panel (d) should be chosen instead. This value fails to detect the individual vortices, instead grouping them in pairs. It does, however, detects the meandering jet as an individual coherent structure. Looking at the maximum eigengap, the value of  $r = 1.5$  in panel (c) should be chosen.

## 2.3 Updated spectral clustering approach with noise perturbation metrics

The spectral clustering approach described above has one less input parameter than FCM or K-Means since the number  $K$  of clusters to detect is determined automatically. The clustering results are also easily repeatable for the same initial parameter settings, in contrast to FCM or K-Means that rely on a random initial distribution of probabilities. Spectral clustering still depends on the numerical offset parameter and the radius  $r$  of sparsification. Moreover, the choice of distance function also influences the eigengap. With the goal to minimize the number of user-input parameters needed, a modified spectral clustering protocol is proposed.



### 2.3.1 Deviations from Hadjighasem *et al.* [2016]

The size and shape of clusters change with the radius  $r$ , as was covered in the previous section. HA16 suggest “the existence of a critical radius” above which they converge. From parameter sweeps on several benchmark flows, including the Bickley Jet example, however, there is not always convergence of the spectral decomposition with increasing  $r$ . Instead, this thesis argues that the  $r$  values are physically relevant and dictate the size of the clusters that are detected, which was illustrated in figures 2-6 and 2-8. Furthermore, the parameter  $r$ , along its corresponding spectral decomposition and associated number  $k$ , should be chosen according to the eigengap heuristics, picking the  $r$  value that maximizes the eigengap:  $k$  is the number of eigenvalues preceding the maximum eigengap; the number of the first  $k$  eigenvectors corresponds to the number of connected components within the graph. The  $r$  value, or values, should be chosen to yield local maxima, or peaks, in the eigengap. In this thesis, the normalized eigengap will be considered. The reason why peaks are considered instead of absolute maxima is because the protocol seeks the structures that are well intra-connected in comparison to structures of similar sizes. Figure 2-6 showed how for  $r$  large enough, the resulting clusters can enclose a large part of the domain; in figure 2-8, the absolute eigengap is greater for  $r = 4.5$  than for  $r = 3.0$ , but it is debatable whether this partition brings more useful information about the flow. A cluster that englobes, say, half the domain will have a high intra-similarity and high inter-similarity, leading to a high eigengap, but the results are of little interest.

In HA16, the offset value for the diagonal of elements  $w_{i,i}$  in  $W$  is proven to be “immaterial”, analytically. In practice, however, the scalar chosen for this offset also impacts the spectral clustering results significantly. As a reminder, the purpose of the offset is to assign the highest possible weight of self-similarity to the nodes  $x_i$ ’s: compared to the weights  $w_{i,j}$  of the edges  $e_{i,j}$  of neighboring nodes  $x_j$ ,  $w_{i,i}$  should be orders of magnitude larger. Very large values (towards  $10^{16}$ ) can introduce numerical errors due to the limits of double-precision. On the opposite, an offset value that is too small compared to the other  $w_{ij}$ ’s entries of  $W$

can yield erroneous results. Unlike  $r$ , the offset is not a real physical parameter, but rather a numerical manipulation to avoid infinity in the diagonal. A convergence in the spectral decomposition is expected for large values of the offset, as long as these scalar values are numerically reasonable. ‘Large’ values, however, will depend on the flow: for this reason, the approach proposed in this thesis sets the offset value as a coefficient related to the maximum  $w_{ij}$  entry in  $W$ . For instance, the offset can be set to be  $\max(w_{ij}) \times 10^n$ . First, this power of offset coefficient is fixed to look at the  $r$  sweep of eigengaps, then the power  $n$  is increased until convergence in the spectral results is reached.

Another deviation from HA16 concerns the amount of sparsification, which is defined by the percentage of nodes  $x_i$  that are kept within the graph  $V$  after the  $r$  threshold is applied. The authors choose  $r$  “such that only 5%–10% of  $W$  is kept” as a rule of thumb, or, equivalent, such that the matrix  $W$  is sparsified at 90–95%. In practice, however, the  $r$  values chosen in different examples from the literature are such that the range of sparsification varies between 88% and 99%. Furthermore, additional parameter exploration seems to indicate that the percent of sparsification of  $W$  is very subjective. For example, in producing the results of figure 2-8, it is shown that the value of  $r = 4.5$  was the only one that left  $W$  95% sparsified; all other values, including the value that was chosen for HA16, yielded sparsification percentages higher than 98%. Looking at the maximum eigengap (“MaxGap” in the figure), however, reveals that the value of  $r = 4.5$  did not yield the optimal partition of the flow. As was defined in section 2.1, the optimal partition is one where the clusters are the most coherent in comparison to the incoherent background. It is therefore argued here that the sparsification percentage cannot be used to determine the optimal  $r$  and that the value of the eigengap should be the only deciding factor in choosing the correct partition of the domain.

The absolute value of the eigengap, as well as the magnitude of the eigenvalues of  $L$  vary with  $r$ ,  $w_{i,i}$  and the offset. Therefore, for convergence of the spectral results, the shape of the spectral decomposition is considered: we look for relative, normalized eigengap as opposed to an absolute value. This relative eigengap, which we refer to as *Gap Ratio*, is the absolute

value of the maximum eigengap over the range from the minimum eigenvalue  $\lambda_{min}$  to the maximum eigenvalue  $\lambda_{max}$ :

$$\text{Gap ratio} = \frac{|\text{maximum eigengap}|}{(\lambda_{max} - \lambda_{min})} \quad (2.10)$$

An illustration is provided in figure 2-9 on page 69. Convergence of the gap ratio means that the shape of the eigenvalue plot is preserved. As was the case with the absolute eigengap, the highest gap ratio means that the intra-subset similarity is the highest and the inter-subset similarity is the lowest. Since the whole argument for spectral clustering relies on the eigengap heuristic, looking at the  $r$  values that yield the local maxima, or peaks, in gap ratios ensures that the domain is partitioned as optimally as possible.

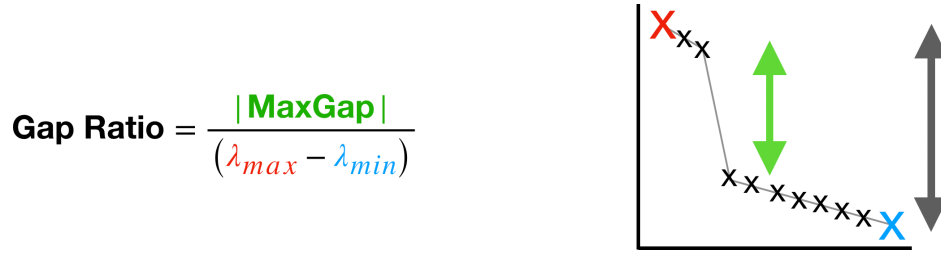


Figure 2-9: The gap ratio from equation (2.10) for the plot of eigenvalues  $\lambda_i$ , plotted with crosses. The gap ratio is the ratio of the maximum spectral gap MaxGap, shown with the small green arrow, over the range of eigenvalues, shown with the big gray arrow. This range is the difference between  $\lambda_{max}$ , in red, and  $\lambda_{min}$ , in blue.

The physical relevance of the radius  $r$ , which should be maximizing the gap ratio, on the length scale of the detected structures was illustrated in figure 2-6. The size of the detected vortices increased gradually from  $r = 0.5$  to  $r = 5.0$ . For  $r \geq 4.0$ , the meandering jet is detected and for  $r > 4.0$ , the vortices are clumped together into pairs. This trend was observed for the majority of the flows: as  $r$  increases, the size of the detected structures increases, until the structure encloses the majority of the domain, showing a high value for connectivity and, correspondingly, a high gap ratio. Indeed, if half the domain is comprised within a single structure, this subset will show high connectivity relatively to the rest of the domain. This result is trivial and not of interest of the analysis of the flow. At very small

values of  $r$ , the numerical resolution is such that the number of nodes found within the  $r$  radius is very low, which can result in computations on subsets having a new nodes only: these results are erroneous. While the gap ratios can be high for both small and large values of  $r$ , they are not peaks; peaks, rather, indicate that structures at this  $r$  value are much more connected in comparison to the incoherent background than the structures at neighboring values of  $r$ . These are the results of interest that are investigated in this protocol.

To summarize, the proposed approach sweeps through  $r$  to pick the values corresponding to local maxima, or peaks, in gap ratio, with the offset value  $w_{i,i}$  corresponding to a coefficient of the maximum  $w_{i,i=j}, i \neq j$ , at each  $r$ . Then, for each  $r$  corresponding to a peak, the protocol sweeps through the coefficients  $w_{i,i}$  to reach convergence in gap ratios. The proposed protocol uses the average distance function for its calculations. Not only are the gap ratios often higher for the average function than for the maximum function, but the outside filamentation that is often included in the clusters detected with the average function can provide very valuable information. With the offset coefficients  $w_{i,i}$  and  $r$  determined through this normalized gap ratio, as well as the number of clusters related to the number of eigenvalues before the eigengap, the new protocol becomes parameter-free.

### 2.3.2 Coherence metrics

Spectral clustering will partition the domain into coherent clusters and an incoherent mixing region. Once the coherent clusters have been identified, there is no standard means of knowing how coherent they are relative to each other or relative to the background. The eigenvalues of  $L$  cannot be used to quantify the coherence of the output clusters because the  $k$  eigenvectors of  $L$  do not generally correspond to the  $K = k + 1$  identified clusters. To determine how coherent the detected clusters are, we define a metrics for coherence: the robustness of the cluster to random noise perturbation. This is similar, in spirit, to the approach by Froyland [2013], who investigated the role of noise amplitude on coherent sets.

The robustness to noise is computed as follows. For a trajectory set initially located

inside a given cluster and advected over  $[t_0; t_f]$ , take the final positions at time  $t_f$ . Estimate the average distance between particles  $D$  and add random noise between  $[0; D/100]$  to these final positions. Advect the ‘perturbed’ positions in backward time to the initial time  $t_0$ . Then count how many perturbed particles fall within the original clusters’ boundaries. The percentage of particles that will return within the boundaries of their assigned cluster is the proposed coherence metrics.

An illustration of the proposed procedure is provided in figure 2-10 on 72. For a flow within a domain  $L_x$ -by- $L_y$  over time interval  $[t_0; t_f]$ , carry-on the spectral clustering protocol by, first, sweeping through values of  $r$  while keeping the offset fixed; second, for each of the local maxima in gap ratios, verifying convergence in the offset coefficient. The example assumes  $K = 3$  clusters identified above and below a jet-like feature and a corresponding eigengraph showing three eigenvalues before the maximum eigengap. The clusters are shown in green, pink and turquoise. The coherence metrics is computed for the three clusters and the fourth incoherent background cluster. The positions at  $t_f$  are slightly perturbed, then advected backward to  $t_0$ . For each of the clusters, the percentage of particles that return within the original boundaries extracted at  $t_0$  is computed. In this example, the coherence values were 100% for the green cluster, 87.5% for the pink cluster, 90% for the turquoise cluster and 90% for the incoherent background. Note that the incoherent background has a coherence percentage similar to the coherent clusters: it is a direct measure of the number of particles that return within the boundaries. The purpose of the metrics is to compare the detected coherent clusters: for example, to evaluate the coherence of the vortices in the Bickley jet example. Generally, with this coherence metrics, large sets tend to be more coherent because of the large volume-to-perimeter ratio.

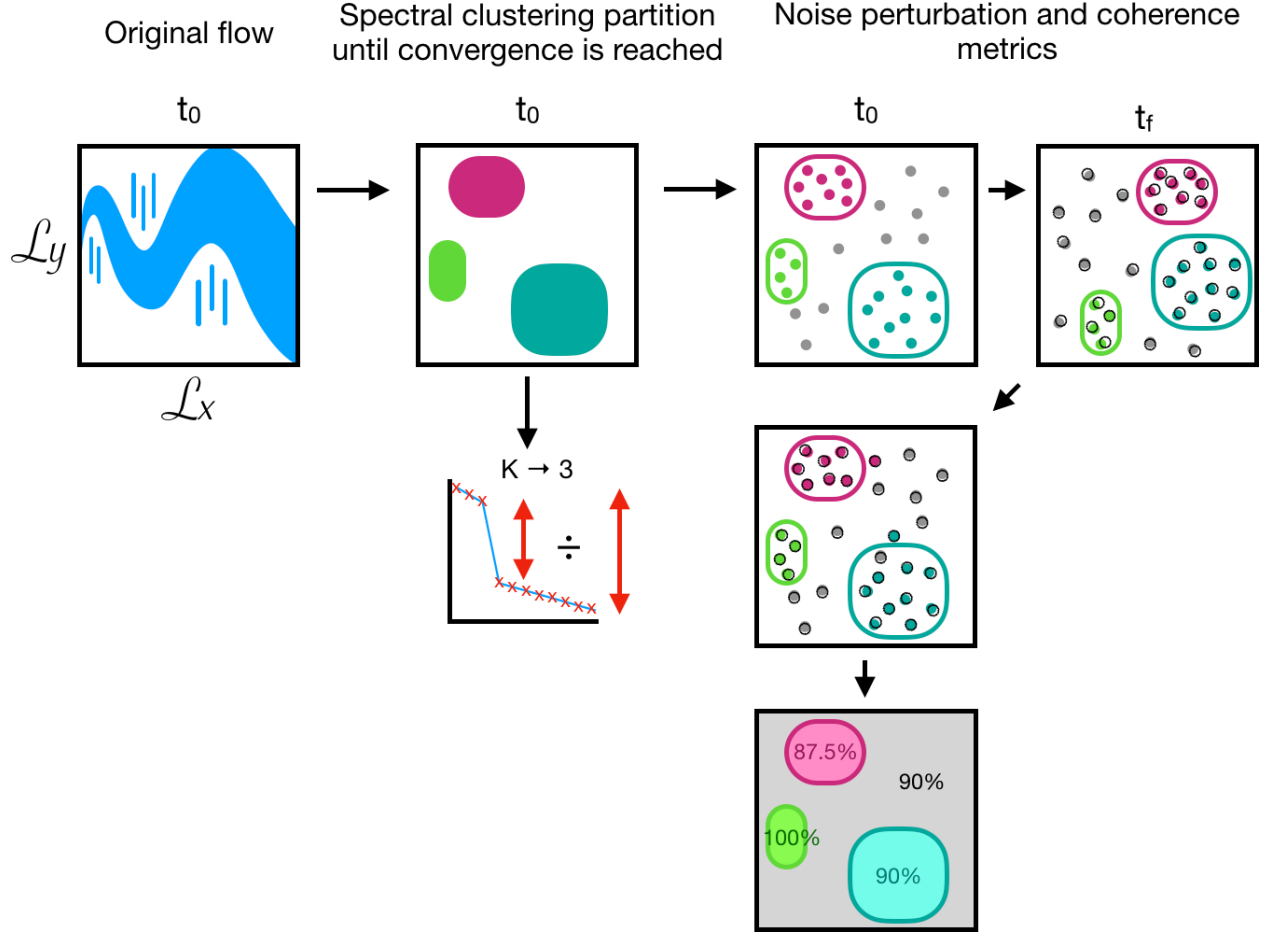


Figure 2-10: Illustration of the spectral clustering protocol with coherence metrics. For a toy model flow from  $t_0$  to  $t_f$ , the parameter-free spectral clustering protocol sweeps through values of  $r$ , with  $r < L_x$  and  $r < L_y$ , and through the power of offset coefficients until convergence in the gap ratio, representative of the spectral decomposition. Here, for example, there are  $k = 3$  eigenvalues before the eigengap, which means 3 connected components detected. For the noise metrics, the particle positions at the final position  $t_f$  are perturbed with random noise, then advected back to  $t_0$ . The percentage of particles that fall back within the clusters boundaries at  $t_0$  is the coherence metrics. This is done for all 4 clusters: the 3 connected components (in colors) and the incoherent background (in gray).

A key component of the coherence metric calculations is finding the boundaries of the clusters. Extracting boundaries is a notoriously hard problem to implement numerically and it becomes especially arduous when clusters have filaments or lobes. The numerical resolution will typically be the limitation for the extraction of appropriate boundaries around the clusters, so outliers at the extremities of the lobes or filaments may need to be removed. Keeping in mind the purpose of the protocol, which is to carry on an objective clustering

analysis for LCS detection with as few user-input parameters as possible, boundaries must be extracted in an automatic manner without user input.

One solution to this problem of filamenting boundaries is to automatically remove these distant outliers by filtering particles that are too distant from the bulk of the cluster. To proceed, calculate the mean pairwise distances for all particles within a cluster and remove particles that are statistical outliers, such as, for example, particles that are more than 4 standard deviations above the mean. Figure 2-11 on 74 shows an example for two Bickley jet vortices and for the a cluster in the asymmetric Duffing oscillator, described in section 2.4.3. For the vortex example with  $r = 0.5$ , the area to perimeter ratio is much smaller than for  $r = 2.0$ , as the detected vortices are much smaller and the number of numerical nodes within the boundaries are much fewer. Therefore, even small changes in the number of particles will affect the results. Applying the  $4\sigma$  threshold to the mean pairwise distances yields appropriate boundaries in both cases. The filaments detected with  $r = 2.0$  are still included in the boundary. The last example is a very filamented cluster, in figure 2-11c; nevertheless, the procedure using the  $4\sigma$  filtering extracted boundaries that were faithfully following the filaments.

### 2.3.3 Algorithm Summary

1. Sweep to determine the  $r$  value(s) with the largest gap ratio
  - For the matrix  $W$  with weights  $w_{ij}$ , set the diagonal offset coefficient to a fixed value  $\max(w_{ij}) \times 10^n$ , with  $n$  arbitrarily large. Setting  $n = 7$  typically works well.
  - Keeping in mind the domain's length scales, sweep through a series of  $r$ 's, bounding these  $r$  values by, at the lower limit, a value large enough to contain at least a dozen nodes, to, at the higher limit, a value smaller than the domain size.
  - Plot the gap ratio  $\frac{|\text{Max Gap}|}{\lambda_{\max} - \lambda_{\min}}$  as a function of  $r$  and find all local maxima.
2. Sweep of offset coefficients  $\max(w_{ij}) \times 10^n$  to verify convergence in the gap ratio

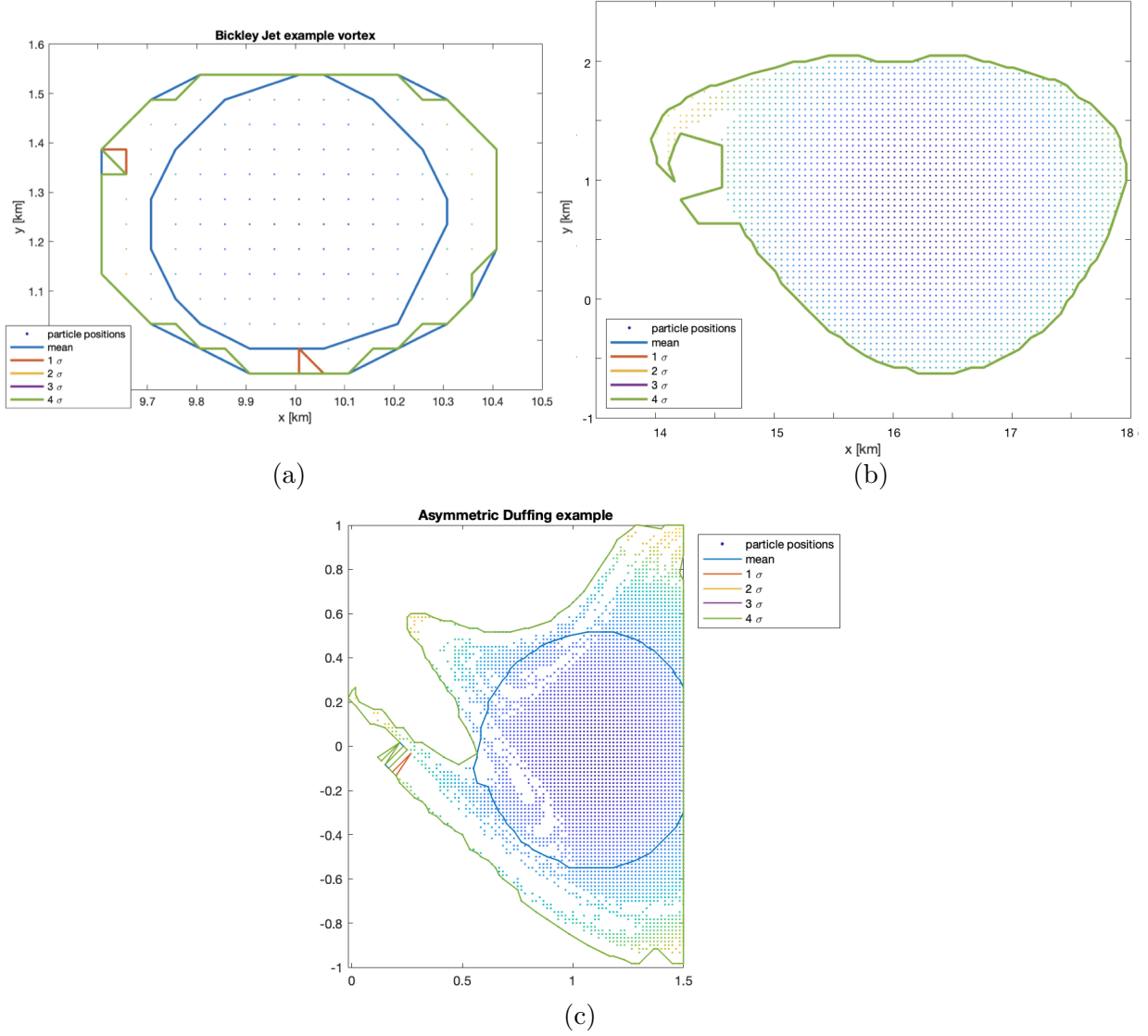


Figure 2-11: Mean pairwise distances between particles and boundary filtering examples. (a) Example vortex for the Bickley jet detected by spectral clustering with  $r = 0.5$ . (b) Same example for a cluster detected with  $r = 2.0$ . (c) Example cluster detected in the asymmetric Duffing oscillator, described in section 2.4.3.



- For all local maxima, vary  $n$  in the offset coefficient  $\max(w_{ij}) \times 10^n$  until both conditions are satisfied (typically, we see convergence for  $n \in [2, 10]$ ):
  - Convergence of the number of detected sets  $k$ , equal to the number of eigenvalues before the gap.
  - Convergence of gap ratio  $\frac{|\text{Max Gap}|}{\lambda_{\max} - \lambda_{\min}}$
- If possible, vary the numerical resolution until the gap ratios converge.

### 3. Noise-based coherence metrics for the resulting clusters

- Take the final positions of particles at time  $t_f$  and add a bit of random noise with amplitude up to  $D/100$ .
- Advect the perturbed positions in backward time to initial time  $t_0$ .
- For each cluster, count how many particles return within the cluster's original boundaries. The percentage is the coherence metrics.

## 2.4 Application to benchmark flows

The protocol described in section 2.3 is applied to three idealized model flows: the Bickley Jet, the Duffing oscillator and the asymmetric Duffing oscillator. Following equation (2.4), the average distance function is used. For comparisons, the analyses conducted with the maximum distance function are included in appendix A: for all three flows analyzed below, the average distance function yielded higher gap ratios, suggesting a more optimal partition of the domain into coherent clusters and an incoherent background.

### 2.4.1 Bickley Jet

As described in Rypina [2007], the Bickley jet consists of a zonal jet background flow on which traveling Rossby waves are superimposed. Rossby waves, or planetary waves, are large-scale inertial waves that result from the conservation of potential vorticity under a

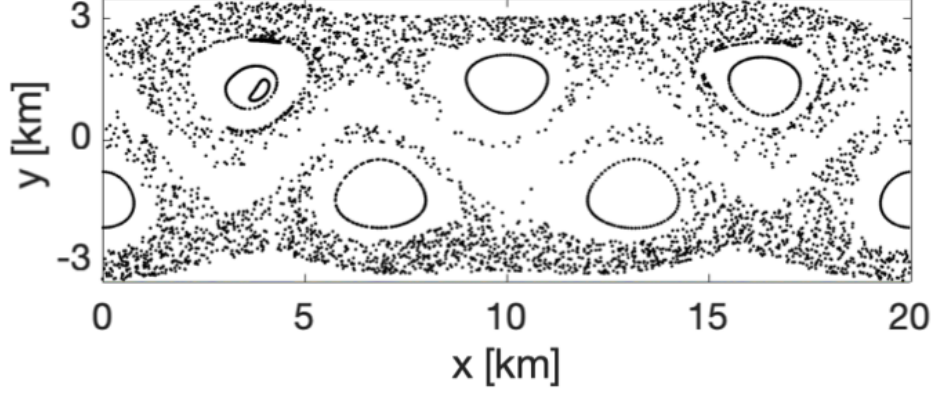


Figure 2-12: Bickley Jet flow reiterated from Rypina *et al.* [2007]. Poincaré map with 1000 periods of integration.

disturbance. In the reference frame of one of the waves, the system consists of a steady background flow and a time-periodic perturbation. The flow is taken from Rypina *et al.* [2007]; the streamfunction is of the following form, with  $N = 2$  waves:

$$\psi(x, y, t) = \psi_0(x, y) + \psi_1(x, y, t) \quad \text{with} \quad (2.11)$$

$$\psi_0(y) = -U_0 L \tanh\left(\frac{y}{L}\right); \quad \psi_1(x, y, t) = U_0 L \operatorname{sech}^2\left(\frac{y}{L}\right) \times \sum_{i=1}^N \epsilon_i \cos[k_i(x - c_i t)]$$

where  $U_0$  is a speed of  $62.66 \text{ ms}^{-1}$ ,  $L$  a reference length of 1770 km,  $k_i$ 's are wavenumbers such that  $k_n = \frac{2n}{r_e}$  with  $r_e$  the Earth radius. The phase speeds are  $c_2 = 0.205U_0$  and  $c_3 = 0.461U_0$ ; the amplitudes are  $\epsilon_2 = 0.15$  and  $\epsilon_3 = 0.30$ . In the frame moving at speed  $c_3$ :

$$\psi_r = c_3 y - U_0 L \tanh\left(\frac{y}{L}\right) + U_0 L \operatorname{sech}^2\left(\frac{y}{L}\right) \left[ \epsilon_2 \cos(k_2 x - \sigma_2 t) + \epsilon_3 \cos(k_3 x) \right] \quad (2.12)$$

with  $\sigma_2 = k_2(c_2 - c_3)$ . The Poincaré map for the flow is shown in figure 2-12 on page 76. The zonal jet separates the flow into two chaotic zones that are characterized by the dense stroboscopic mapping of the particle trajectories after a 1,000 periods of perturbation. The jet itself is detectable by the absence of trajectories crossing through meridionally. On each side of the jet, 3 recirculation vortices are spun, as characterized by the closed orbits. These orbits bound islands of coherence and correspond to the boundaries of the vortices.

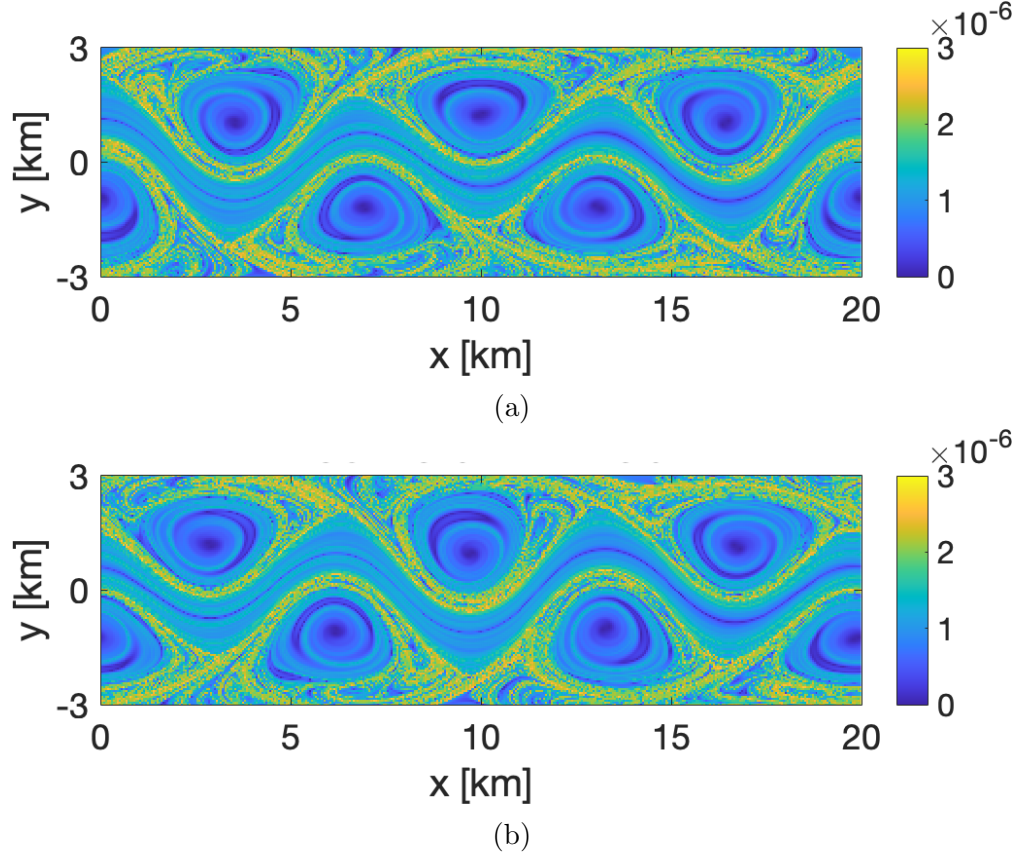


Figure 2-13: Quasi-periodic Bickley Jet flow reiterated from HA16. (a) Forward FTLE field. (b) Backward FTLE field.

The flow used in HA16 was slightly modified with  $N = 3$  waves, making the time dependence of the flow quasi-periodic: the term  $\epsilon_1 \cos(k_1 x - \sigma_1 t)$  is added to equation (2.12), with  $\epsilon_1 = 0.0075$ ,  $\sigma_1 = k_1(c_1 - c_3)$  and  $c_1 = c_3 + (\sqrt{5} - \frac{1}{2}) \times \frac{k_2}{k_1}(c_2 - c_3)$ . The flow is similar to the periodic system, still, and it is expected that the spectral clustering approach would detect the coherent vortices within the flow, as was the case in HA16 and HA17. It is important to remember, however, that the purpose of the spectral clustering protocol is to maximize the intra-cluster similarity and the inter-cluster dissimilarity, so clusters do not need to coincide with the Poincaré map's closed curves, which have the highest intrinsic coherence. The FTLE fields for the quasi-periodic flow were computed according to the method by Farazmand & Haller [2012] and are shown in figure 2-13 on page 77.

Following the proposed parameter-free spectral clustering approach, the first step is to

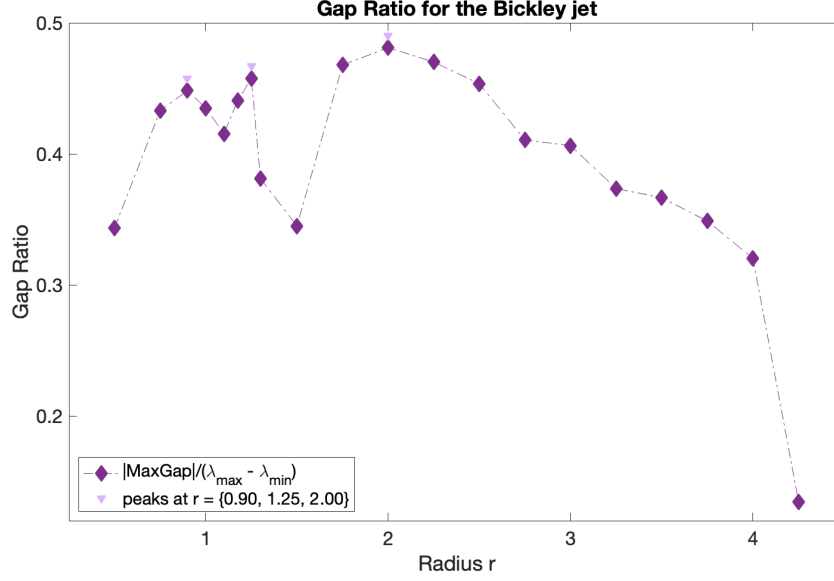


Figure 2-14: Step 1 of the spectral clustering protocol for the Bickley Jet example: gap ratio as a function of  $r$  with the average distance function.

determine the gap ratio dependence on the radius  $r$ . The sweep of  $r$  values is shown in figure 2-14 on page 78 for an initial offset coefficient of  $10^7$ . The average distance function yielded slightly larger gap ratios than the maximum distance function, shown in appendix A. With the average distance function, the peaks in gap ratios were found for  $r = [0.90, 1.25, 2.00]$ . The convergence of parameters for offset coefficients  $10^n$  is verified in all cases, as shown in figure 2-15 on page 79. In all cases, the spectral gap occurs after  $k = 6$  eigenvalues and the eigengap converges at offset value of  $w_{i,i} = 10^7$ .

The clustering results and the coherence metrics were investigated for all peaks, as shown in figure 2-16 on 80. In all cases, the protocol successfully detected the 6 recirculation vortices, annotated by clusters 1 to 6. The clusters were located in the vicinity of the closed orbits in the Poincaré map. The sizes of the clusters varied with the radius  $r$  of sparsification, with very small clusters detected for  $r = 0.9$  and larger clusters for  $r = 2.0$ . The larger clusters contained outside filamentation, as was expected with the average distance function. For  $r = 2.0$ , which corresponded to the highest peak in gap ratio detected, the coherence percentage of the clusters exceeded the percentage of the incoherent background, cluster 7:

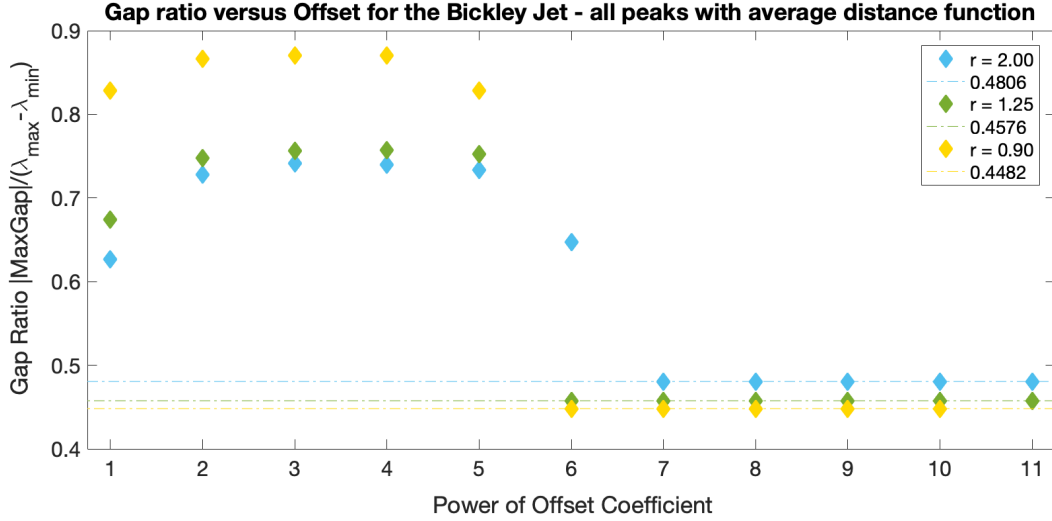


Figure 2-15: Step 2 of the spectral clustering protocol for the Bickley jet: sweep of offset coefficients  $10^n$  for the gap ratio peaks at  $r = 0.90$ ,  $r = 1.25$  and  $r = 2.00$  with the average distance function.

this shows that the optimal partition of the domain according to the gap ratio width is one that maximizes intra-cluster similarity while, at the same time, minimizing the similarity between clusters and with the incoherent cluster.

## 2.4.2 Duffing oscillator

The Duffing system is a flow often used as a benchmark in dynamical system theory, as it exhibits chaotic behavior within a certain parameter regime. It is an oscillator coupled with a cubic stiffness term that produces two gyres with the same sign of rotation, which are symmetric about a saddle point located at the origin. The streamfunction is taken from Rypina & Pratt [2017]:

$$\psi(x, y, t) = [1 - \epsilon \cos(\omega t + \phi)] \left( \frac{x^2}{2} - \frac{ax^4}{4} \right) - \frac{y^2}{2} \quad (2.13)$$

Here,  $\epsilon = 0.1$ ,  $\omega = \frac{3\pi}{2}$ ,  $\phi = \frac{\pi}{4}$  and  $a = 0.5$ . The corresponding Poincaré map and FTLE fields are shown in figure 2-17 on page 81. The two gyres are contoured by closed orbits in Poincaré maps and by low FTLE values. The surrounding domain has a very strong chaotic behavior,

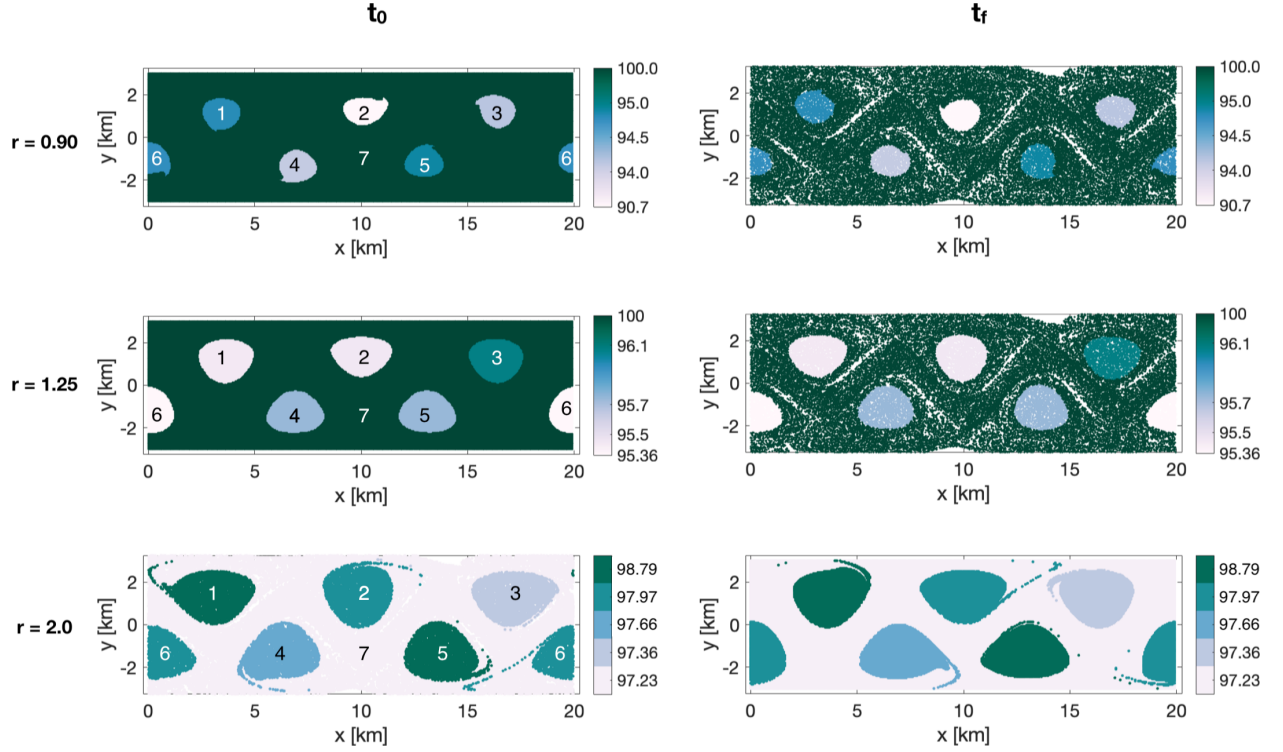


Figure 2-16: Step 3 of the spectral clustering protocol for the Bickley jet: results with coherence metrics for all peak gap ratios with the average distance function. All 7 clusters are annotated. The peaks were detected at  $r = 0.90$ ,  $r = 1.25$  and  $r = 2.0$  with the average distance function. The initial ( $t_0$ ) and final ( $t_f$ ) positions are plotted.

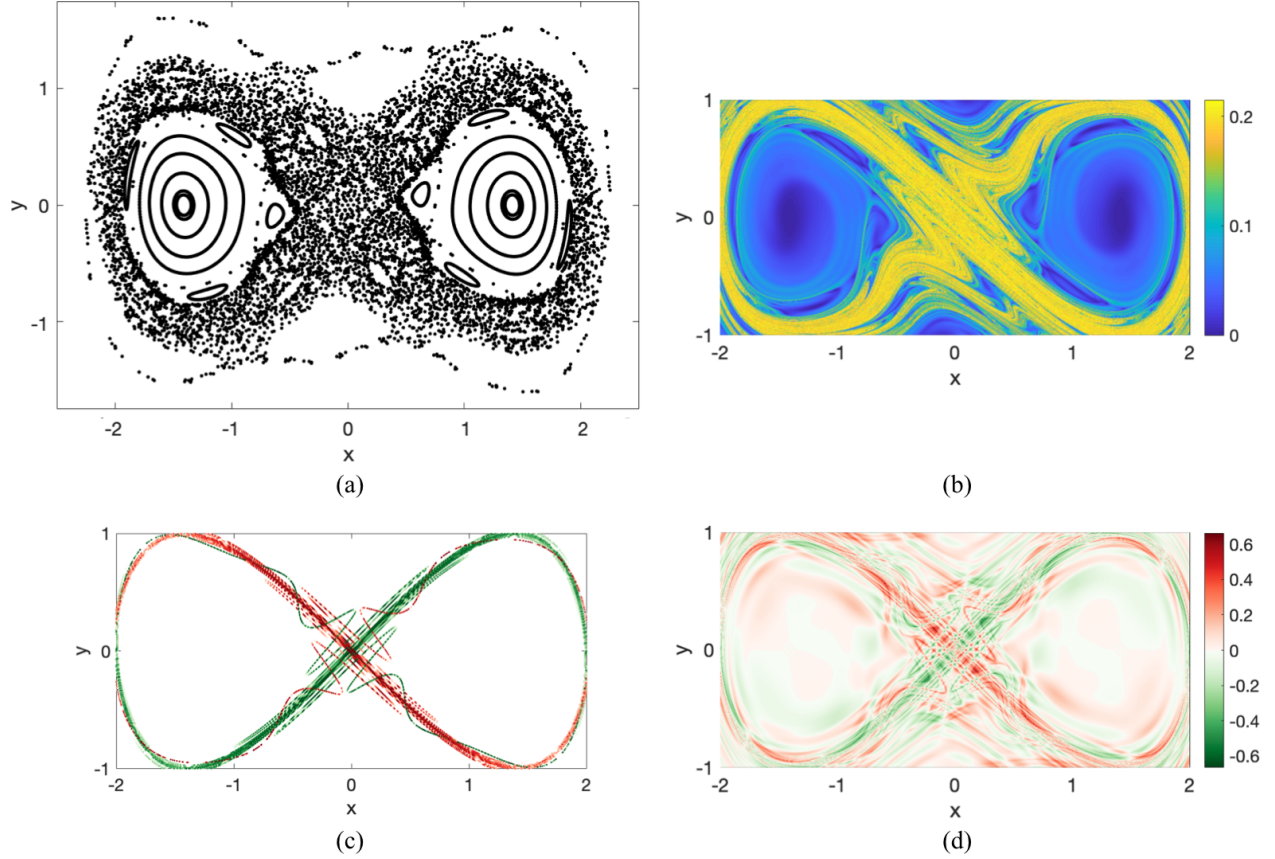


Figure 2-17: The Duffing oscillator. (a) Poincaré map with 1000 periods of perturbation  $T_{pert} = 2\pi/\omega$ . (b) Forward time FTLE for  $30T_{pert}$ , time chosen for the spectral clustering analysis. (c) FTLE ridges in forward (red) and backward (green) time for  $10T_{pert}$ . (d) Superimposed forward (positive, red tones) and backward (negative, green tones) FTLE fields for  $10T_{pert}$ .

as shown by the dense stroboscopic mapping after 1,000 periods of perturbation  $T_{pert} = \frac{2\pi}{\omega} = \frac{4}{3}$  and larger FTLE values. The plot of FTLE ridges shows the classic homoclinic tangle intersection of stable and unstable manifolds and the lobe behavior around the saddle point. Thus, for this flow, we expect two coherent clusters and a figure-eight-shaped incoherent background.

The parameter-free spectral clustering protocol was followed. The first step, the sweep in  $r$ , revealed two peaks in gap ratios at  $r = [1.0, 2.0]$  for the average distance function. The sweeps of parameters are shown in figure 2-18.(a) on page 83, followed by the second step of the protocol, the convergence verifications for the offset coefficients, shown in figure 2-18.(b).

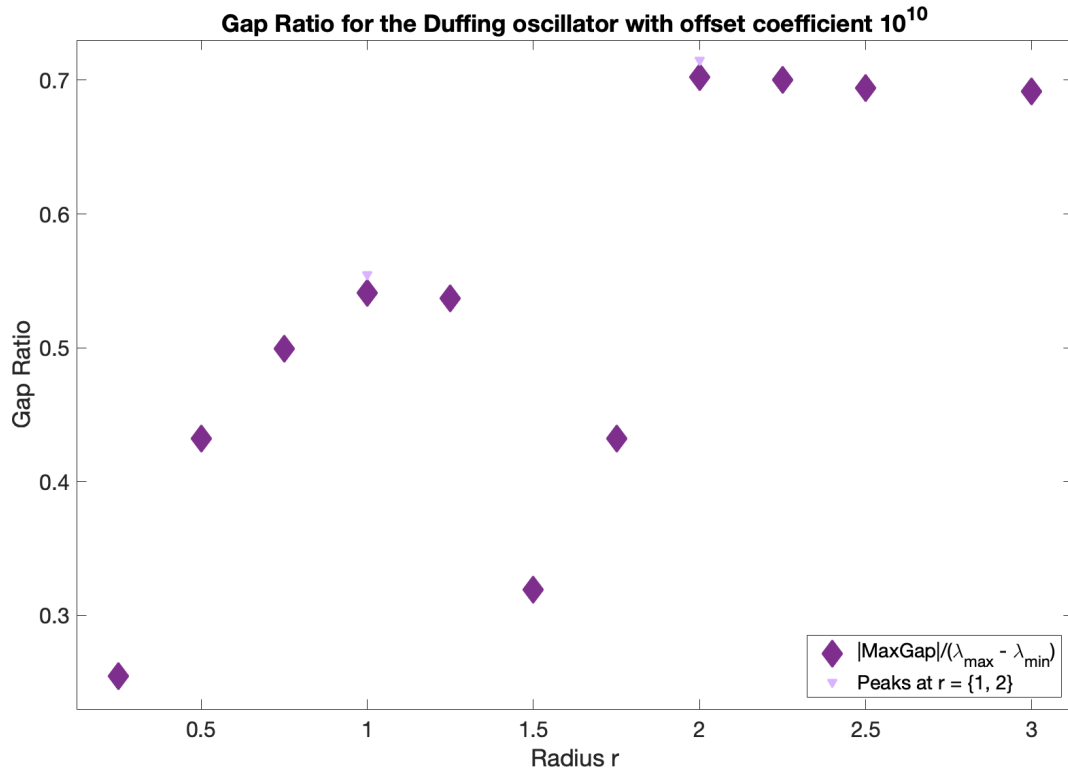


For all cases, convergence in gap ratios was attained after offset coefficient  $10^7$ . The average distance function yielded larger gap ratios on average. The final step of the protocol, the coherence metrics results for each cluster, is shown in figure 2-19 on page 84.

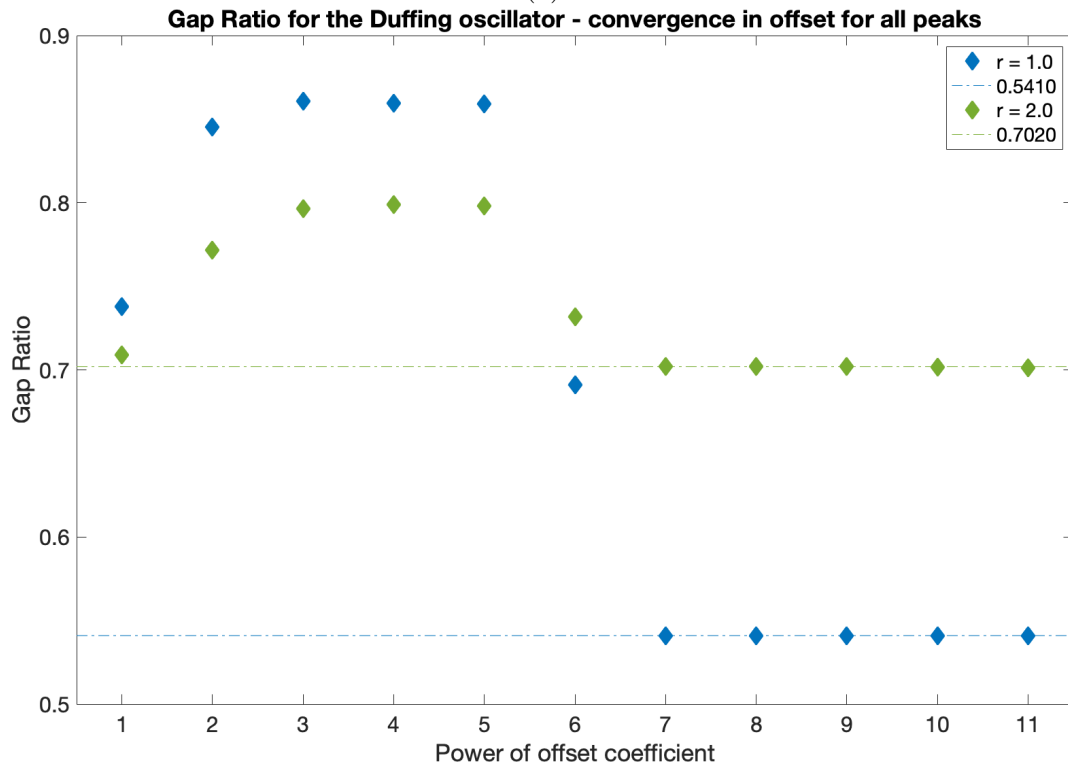
For the first peak detected with the average distance function, two clusters were detected, one around each gyre: clusters 1 and 2. The gap ratio corresponded to  $r = 1.0$  and was higher than the gap ratios from the maximum distance function, shown in appendix A. The spectral clustering approach indicates that the optimal partition of the domain yields clusters that included filaments around the gyres. Looking at the coherence metrics of the clusters and of the incoherent background, cluster 3, the coherence-to-incoherence ratio increased with the gap ratio: for this last peak at  $r = 1.0$ , the clusters had a higher coherence value (above 96.38 %) than the incoherent background, with a coherence percentage of 89.56. This incoherent background had a lower percentage than the incoherent background from the maximum function results. Furthermore, superimposing the FTLE ridges to these results, the clusters at initial time are delimited by the forward-time FTLE ridge, a repelling hyperbolic LCS, and at final time, after 30 periods of perturbation, they are delimited by the backward-time FTLE ridge, an attracting hyperbolic LCS. The spectral clustering analysis was conducted over  $30T_{pert}$  but the FTLE ridges were computed for  $10T_{pert}$  for clarity. This indicates that the filaments included within the clusters are the lobes. This backs the claim that spectral clustering partitions the domain with highest intra-cluster similarity and lowest inter-cluster similarity. The gyres had similar coherence in both cases, with a difference of only 0.02%.

The larger peak at  $r = 2.0$  corresponded to one cluster (cluster 1) that englobed both gyres. The coherence value for this cluster was of 98.6%, compared to 99.9% for the incoherent background, cluster 2. The boundary calculations were still conducted following the protocol with a filtered boundary around both gyres. The boundaries used for these calculations are shown in figure 2-20 on page 84. The filtering used the average pairwise distance and kept particles within  $4\sigma$  from the mean, as explained in the above section. The proposed filtering extracted very accurate set boundaries, even with a boundary encircling





(a)



(b)

Figure 2-18: Steps 1 and 2 of the spectral clustering protocol for the Duffing oscillator with average distance function. (a) Sweep of  $r$  parameters with offset coefficient  $10^{10}$  and the average distance function. (b) Sweep of offset coefficients  $10^n$  for the gap ratio peaks in (a).

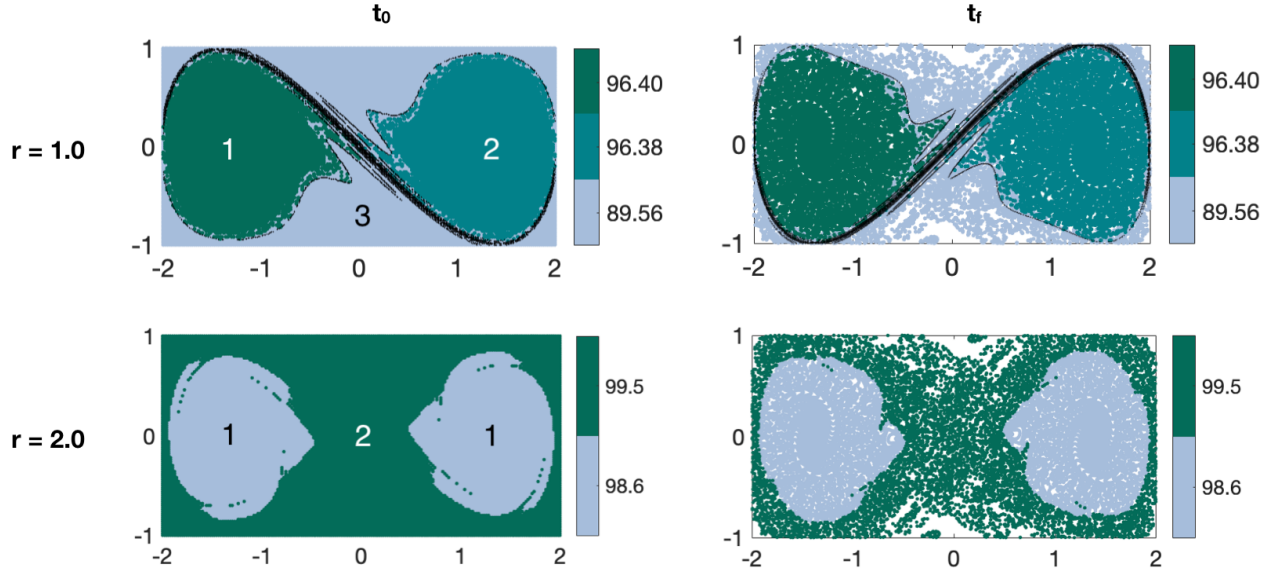


Figure 2-19: Step 3 of the spectral clustering protocol for the Duffing oscillator with the average distance function: results with coherence metrics under noise perturbation and FTLE ridges superimposed. All clusters are annotated. The peaks were detected at  $r = \{1.0, 2.0\}$ . The initial ( $t_0$ ) and final ( $t_f$ ) positions are plotted. The FTLE ridges (black lines) are superimposed to clusters 1 and 2 from  $r = 1.0$ : they correspond to the boundaries of the clusters at  $t_0$  for the forward-time ridge and at  $t_f$  for the backward-time ridge.

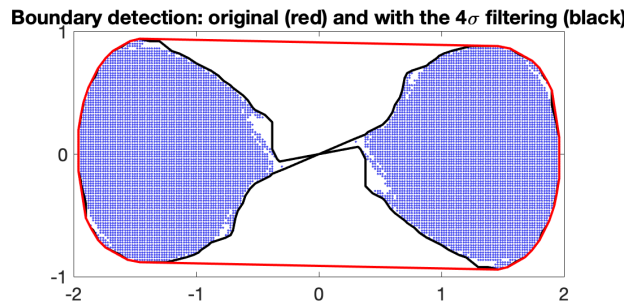


Figure 2-20: Boundary calculations for the single connected set within the Duffing oscillator, detected with  $r = 2.0$ . The protocol was followed step-by-step and extracted the boundaries of the single cluster (containing both gyres) using the  $4\sigma$  filtering.

both gyres within a single set.

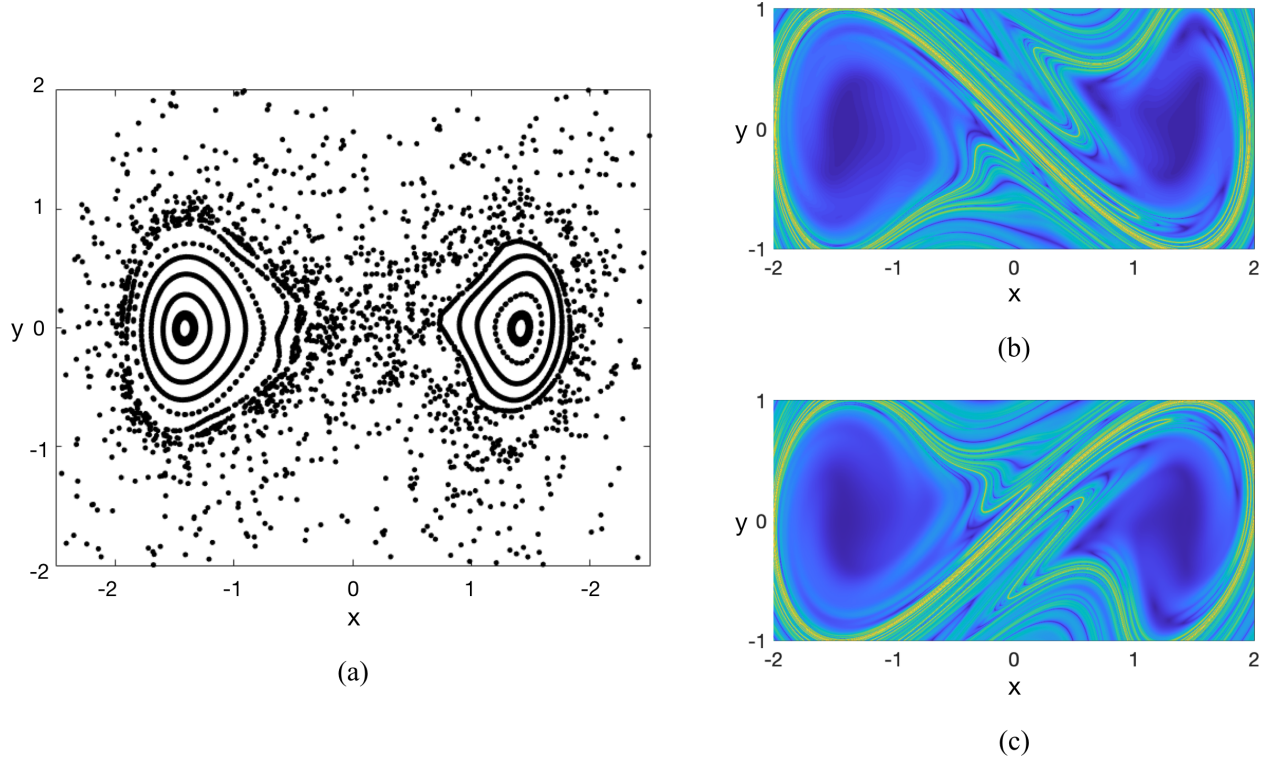


Figure 2-21: The asymmetric Duffing oscillator. (a) Poincaré map with 1000 periods of perturbation  $T_{pert}$ . (b) Forward and (c) backward FTLE for  $10T_{pert}$ .

### 2.4.3 Asymmetric Duffing oscillator

The Duffing oscillator and the Bickley Jet featured vortices of similar sizes. To evaluate the spectral clustering protocol on structures of different scales, an asymmetry was added to the gyres of the Duffing oscillator in equation (2.13):

$$\epsilon \rightarrow \epsilon(x) = 0.1 \times \frac{x + 4.5}{2}$$

The corresponding Poincaré map and FTLE fields are shown in figure 2-21 on page 86. The two gyres are still present but the right gyre is smaller, as contoured by smaller outermost closed orbits. The initial conditions and periods of integration  $30T_{pert}$  were the exact same as for the symmetric Duffing oscillator.

The spectral clustering analysis revealed one peak in gap ratios at  $r = [1.0]$ , as shown in figure 2-22.(a) on page 88, yielding  $k = 2$  eigenvalues before the eigengap. For  $r > 2.0$ , the

spectral decomposition yielded one cluster that englobed both gyres, similar to the symmetric Duffing oscillator. This change in the number  $k$  of connected components explains the shape of the curve. In the symmetric case, a second peak was detected after the jump in gap ratios at  $r = 2.0$ : this corresponded to the cluster enclosing both gyres found with  $r = 1.0$ . Here in the asymmetric case, no second peak was detected, likely due to the differences in scales between the left and the right gyre. For  $r = 1.0$ , convergence in gap ratios was attained after offset coefficient  $10^7$ . The coherence metrics results are shown in figure 2-23 on page 89.

The clusters detected within the asymmetric Duffing oscillator were very filamented. Nevertheless, the procedure using the  $4\sigma$  filtering extracted boundaries that were faithfully following the filaments, as was detailed in figure 2-11c. The right cluster, cluster 2, was much more filamented and had a lower coherence value of 91.87%, compared to 97.0% for the left cluster, cluster 1. Cluster 1 was also larger and the difference in perimeter-to-area ratio also affects the coherence metrics. The incoherent background region, cluster 3, had a coherence value with 85.78%, much lower than either gyre. As was the case for the symmetric system, these clustered filaments corresponded to the FTLE ridges: as shown in figure 2-23, the clusters at initial time  $t_0$  are delimited by the forward-time FTLE ridges and at final time  $t_f$ , by the backward-time FTLE ridges. As was the case for the symmetric Duffing oscillator, the spectral clustering analysis was conducted over  $30T_{pert}$  but the FTLE ridges were computed for  $10T_{pert}$  for clarity. This superimposition suggests that the spectral clustering protocol was able to detect the lobes here as well.

The above examples have shown that the proposed parameter-free spectral clustering approach finds the partition the domain with highest intra-cluster similarity and lowest inter-cluster similarity, with the dissimilarity of the background being maximized. To further demonstrate (1) the ability of the spectral clustering approach to find this partition and (2) the utility of the coherence metrics used in the protocol, the coherence metrics was computed for sets that corresponded to closed orbits of the Poincaré map of the asymmetric Duffing oscillator. Two types of sets were chosen for comparison: a set corresponding to orbits close

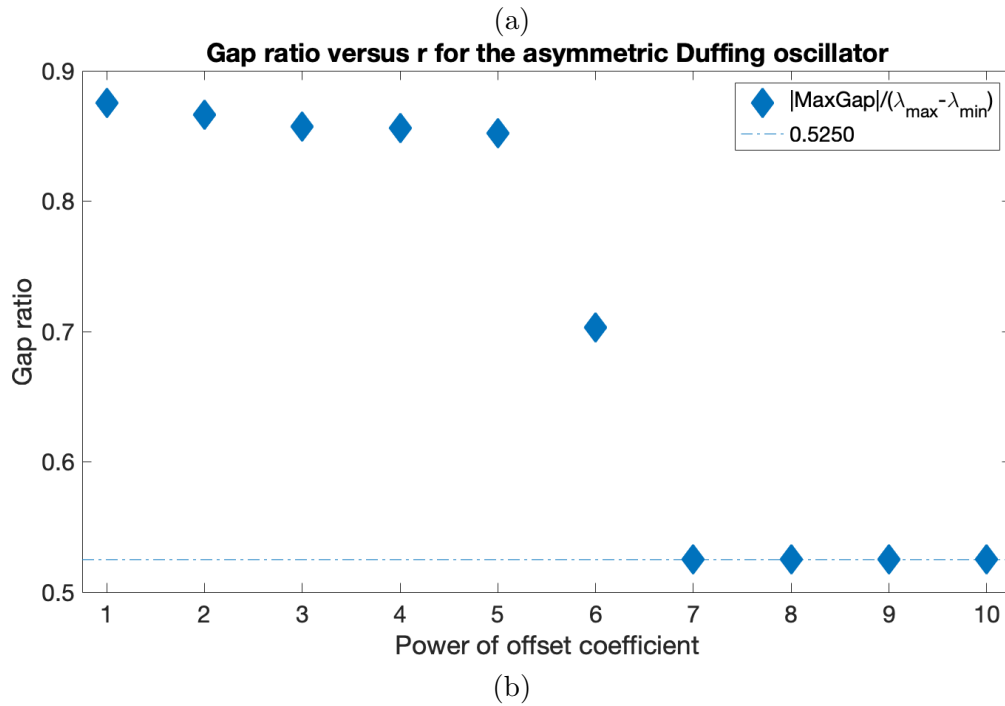
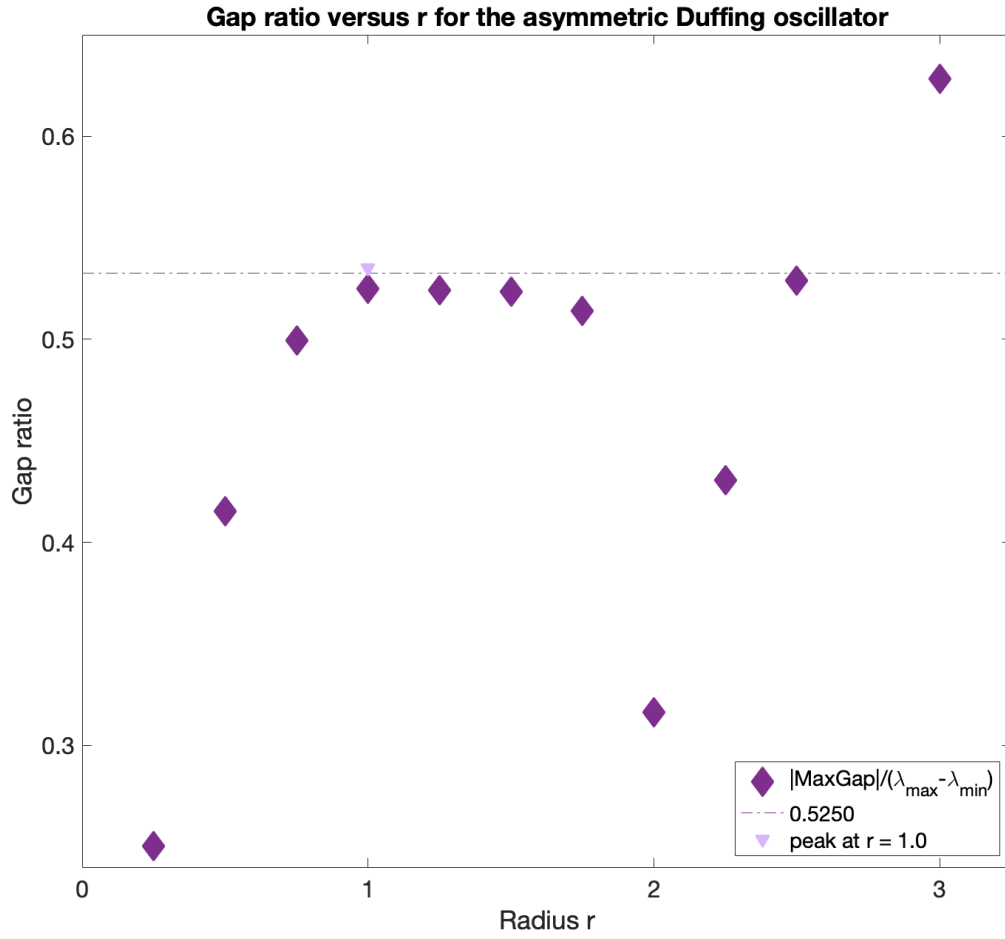


Figure 2-22: Steps 1 and 2 of the spectral clustering protocol for the asymmetric Duffing oscillator. (Top) Sweep of  $r$  parameters with offset coefficient  $10^7$  for the average and maximum distance functions. (Bottom) Sweep of offset coefficients  $10^n$  for average distance function and the gap ratio peak at  $r = 1.0$ .

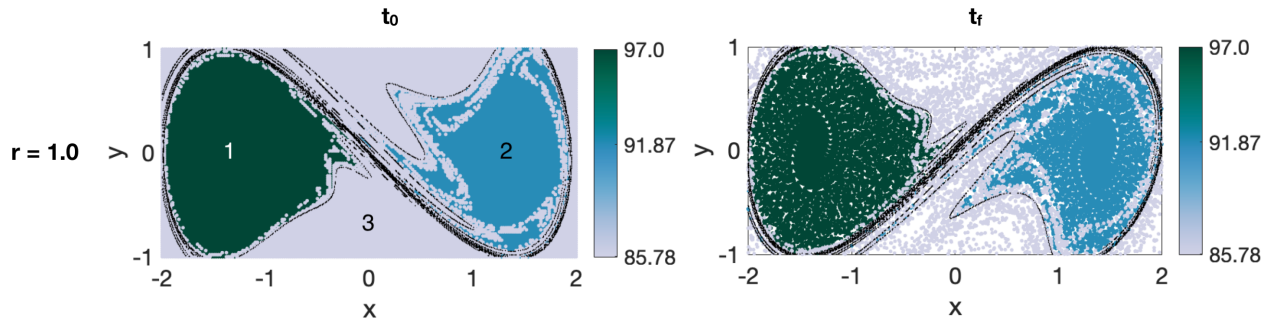


Figure 2-23: Step 3 of the spectral clustering protocol for the asymmetric Duffing oscillator: results with coherence metrics under noise perturbation and FTLE ridges superimposed. Results with the average distance function and peak gap ratio at  $r = 1.0$ : coherence metrics for the clusters at initial ( $t_0$ ) and final ( $t_f$ ) times. The black lines are the FTLE ridges in forward-time at  $t_0$  and in backward time at  $t_f$ .

to the cores of the gyres and a set corresponding to the outermost closed orbits. The results are shown in figure 2-24 on page 90. In each case, the coherence values inside the orbits were higher than the coherence of the detected spectral clusters, but the incoherent region was also much more coherent, with 99.91–100.00 % coherence, as opposed to 85.78% for the incoherent background around the spectral clusters. This suggests that indeed, spectral clustering does not find the structures with the most intrinsic coherence, but rather the partition of the domain that maximizes inter-cluster dissimilarity.

Having successfully tested the proposed parameter-free spectral clustering protocol on 3 benchmark flows, in the next two chapters, the method will be applied to realistic oceanic flows.

## 2.5 Conclusions

Conventional spectral clustering enables the detection of LCS with fewer user-input parameters than other clustering methods, such as FCM or K-Means, but it is not parameter-free. The resulting detected clusters are highly dependent on the radius  $r$  of sparsification and the offset parameter  $w_{i,i}$ . To impartially determine the optimal choice of  $r$  and  $w_{i,i}$ , a modified spectral clustering protocol was proposed.

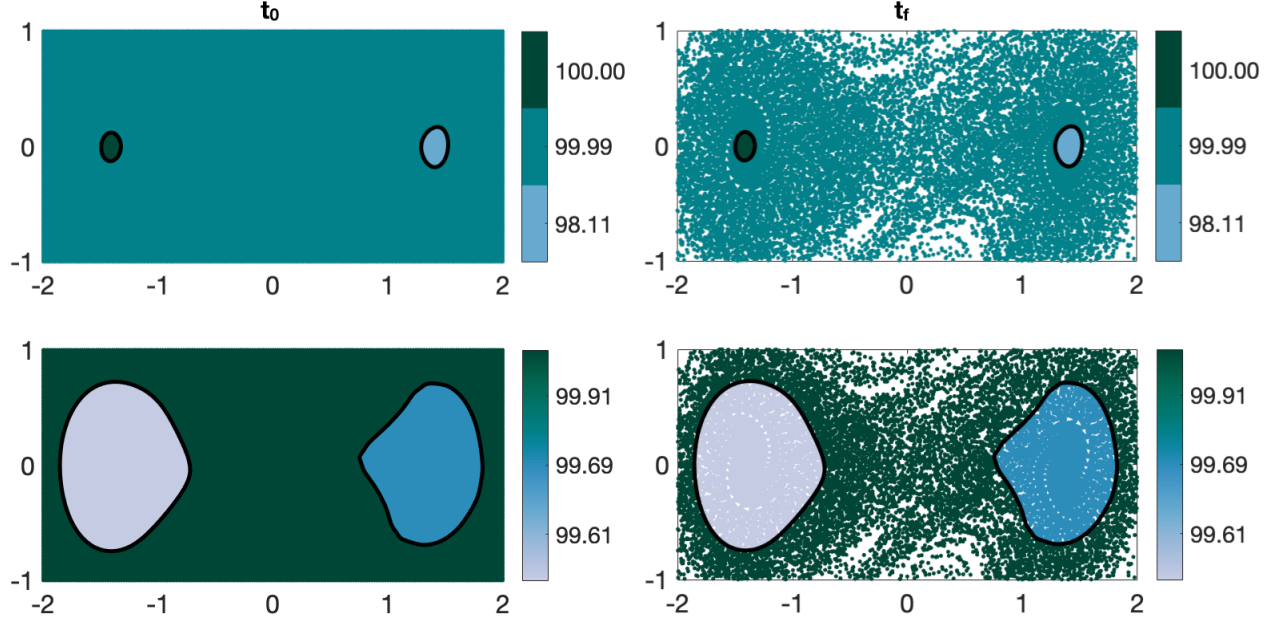


Figure 2-24: Coherence metrics applied to closed orbits (black lines) of the Poincaré map: (Top) Inner orbits close to the gyre core (Bottom) Outer orbits close to the gyre boundaries.

To evaluate the coherence of the resulting clusters, their robustness to noise is computed by slightly and randomly perturbing the trajectories and counting the percentage of trajectories that remain within the boundaries. Filtering the sets by average pairwise distance within  $4\sigma$  from the mean eliminated outliers in filaments and allowed the extraction of very accurate set boundaries.

The accuracy and the usefulness of the proposed protocol and of the coherence metrics have been successfully demonstrated in three idealized model flows. The method will be applied to field studies in the following chapters.



## Chapter 3

# Uncovering transport in a coral atoll with Lagrangian Coherent Structures

To investigate the applicability of the Lagrangian methods discussed in previous chapters to the study of ocean transport, a field experiment was undertaken at Scott Reef, a coral atoll in the Timor Sea. The complex bathymetry and the regular forcing that typically characterize coral reefs make these systems amenable to Lagrangian analyses. Two complimentary LCS methods, the FTLE method and the spectral clustering method, were used for the planning and execution of releases of drifters, which were targeted to identify and evaluate a key, transient flow transport structure predicted to occur at a certain time of the tidal cycle, based on preliminary numerical analyses.

This chapter presents the motivations behind the field experiment in 3.1, then describes the site of study in 3.2, along with the numerical model used in section 3.3. The predictive analysis is described in 3.4. The experimental planning and methods are described in 3.5. The experimental results and further details on the Lagrangian data processing are presented in 5.1 and discussed in 3.7.

### 3.1 Motivations

Coral communities form as a result of the interplay between ocean physics and reef accretion [Monismith, 2007; Lowe & Falter, 2015]. Coral reefs are built by colonies of coral polyps, which secrete calcium carbonate to grow exoskeletons. The aggregation of the deposited calcium carbonate can eventually form an atoll: a coral rim that encircles a lagoon. Typically, the coral rim started as a reef accreting around an island or a volcano, which gradually subsided to erosion while the coral reef kept growing and building upward at a greater rate than the rate of subsidence. Coral reefs are one of the most biodiverse ecosystems on Earth and provide habitat for myriads of marine species [Kahng *et al.*, 2017].

Coral reefs are increasingly subject to environmental stressors associated with anthropogenic activities [Kahng *et al.*, 2010; Hughes *et al.*, 2017; Rivera, 2019]. Within coral reef systems, flow transport plays several important roles, influencing to varying degrees such processes as nutrient exchange, larval settlement, thermal exchange and biological connectivity [Monismith, 2007; Lowe & Falter, 2015; Green *et al.*, 2019]. Understanding the organization of flow transport is thus an important consideration for the study of reef populations and, furthermore, the resilience of these ecosystems [Botsford *et al.*, 2009; Kool *et al.*, 2011]. In Western Australia, for example, strong recent coral reef bleaching events is in part due to the horizontal advective transport of heat [Xu *et al.*, 2018].

In coral sciences, connectivity refers to ‘the process of larval dispersal and subsequent recruitment of individuals into a receiving population’ [Wood *et al.*, 2014]. It is the primary tool to study coral population dynamics as it is used to evaluate the relative importance of different reefs as larval sources and/or as destinations for larval dispersal. The example in figure 3-1 on page 93 is taken from NOAA Coral Reef Watch at <http://coralreefwatch.noaa.gov/satellite/hdf/index.php> and shows the connectivity among coral reefs in the Hawaiian archipelago, determined from tracking simulated larval releases link between different reefs [Kolinski & Cox, 2003; Kool *et al.*, 2010]. This matrix shows that the connectivity between the reefs on the islands of Oahu and Molokai is high, whereas the reefs in Lanai

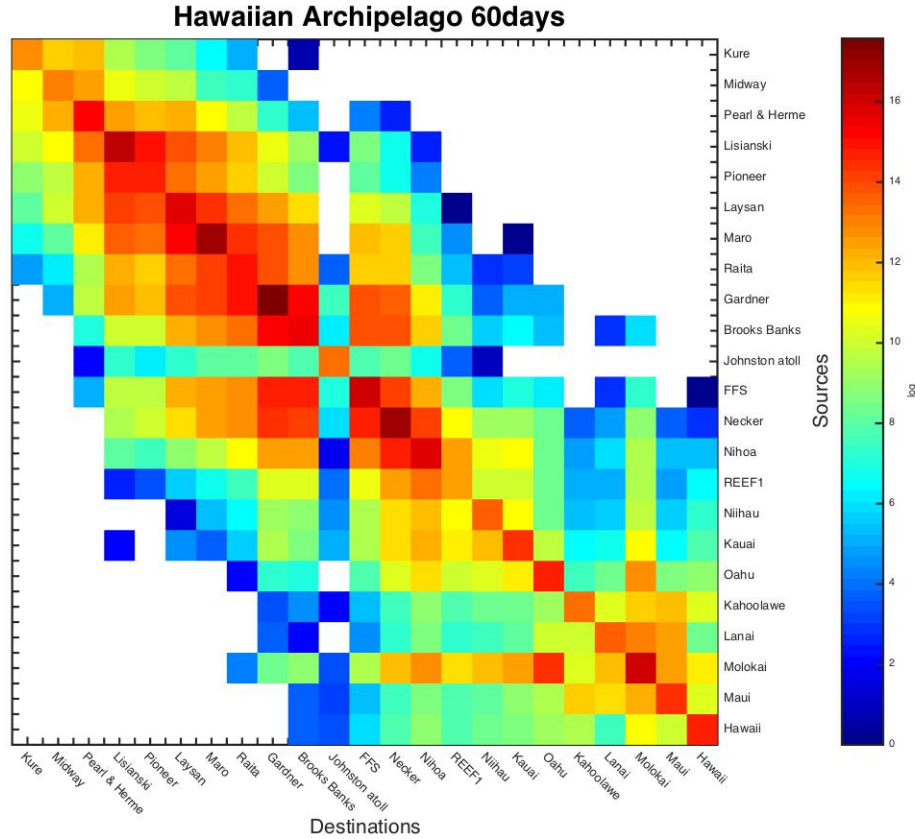


Figure 3-1: Connectivity matrix for the Hawaiian archipelago. Retrieved from NOAA Coral Reef Watch based on the works of Kolinski & Cox [2003]; Kool *et al.* [2010]

seemed more isolated, which may come as a counterintuitive results considering Molokai and Lanai are neighboring islands and Oahu is much further west.

Coral reef systems typically have well defined bathymetry and consistent forcing mechanisms, such as tides or wave breaking at the edge of the reef. The bathymetry tends to shelter the reef and to constrain the length scales of the dynamics of interest within the system. The periodic forcing helps identify the relevant temporal scales. Understanding transport in the spatiotemporally complex flows characteristic of coral reefs is typically challenging, however, as the inherent Lagrangian nature of flow transport makes the identification of transient structures in velocity fields nontrivial. A promising approach is the Lagrangian approach, including LCS detection methods that have been developed to identify the key structures governing Lagrangian transport, as was covered in the previous chapters.

A model study of Leclair *et al.* demonstrated the potential of such Lagrangian approaches for coral reef systems. The authors studied transport in a 16-km long section of Ningaloo Reef at Coral Bay, Western Australia, with the goal to explain habitat connectivity within the reef. As shown in figure 3-2.e on page 95, Leclair *et al.* superimposed the connectivity matrices of the reef on top of the hyperbolic and elliptic LCS detected by the geodesics method, from velocity fields output by a ROMS model. They also combined FTLE fields, the hyperbolic and elliptic LCS (figure 3-2.a-b) and the partition of the domain into subdomains according to the fate of the fluid particles: in figure 3-2.c-d, particles advected towards different cardinal directions are shaded in purple or green shades and particles being trapped inside eddies are shaded in yellow. The maps highlighted how the key transport structures governed reef connectivity and revealed the presence of hyperbolic, separating ridges between neighboring reefs, as well as elliptic LCS forming barriers to transport between the reefs they encircled and other reefs in the vicinity. The presence of such transport barriers explained why geographically close locations may have very low connectivity, whereas locations that are further away may have larval connectivity due to the nature of flow transport. The LCS approach helped understand the data presented in the connectivity matrices.

Connectivity is an important factor in coral resilience as it can allow different reefs to recover from disturbances such as bleaching or reef destruction. Such connectivity tools are heavily taken into consideration during the assessments and planning of conservation efforts, such as the establishment of Marine Protected Areas (MPAs) [Kough *et al.*, 2018]. The Lagrangian approach used in the aforementioned numerical study remained to be demonstrated in the field. Moreover, the FTLE and spectral clustering methods have not been previously used in the planning and execution of field experiments. To evaluate the utility and applicability of these methods for coral reefs, a field deployment of drifters was designed and conducted in October 2016 at Scott Reef.

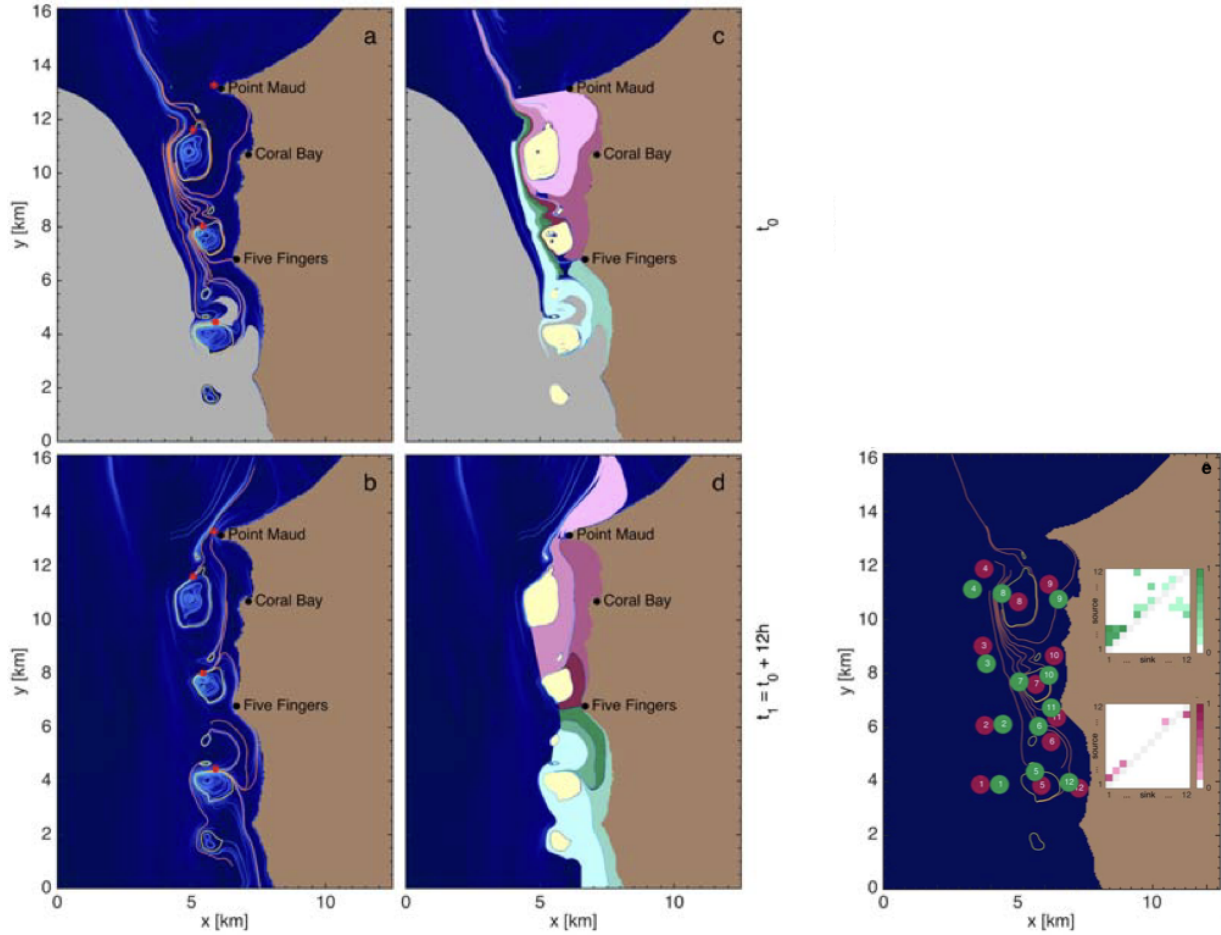


Figure 3-2: Connectivity and LCS in Coral Bay, Australia. Figures and captions extracted from Leclair et al. (a) The forward (repelling; at time  $t_0$ ) and (b) backward (attracting; at time  $t_1$ ) FLTE fields. Plotted are the hyperbolic (orange) and elliptic (yellow) LCS; the hyperbolic are either (a) repelling LCS at the initial time  $t_0$  or (b) the attracting LCS at the final time  $t_1$ , while the elliptic LCS mark the boundaries of Lagrangian eddies at both initial and final times. (c,d) Color-coded particle patches initially distributed in the regions defined by repelling LCS are presented in (c), and their final distribution is presented in (d). (e) LCS and connectivity matrices. The locations of two slightly different ( $< 300$  m offset) release case studies (purple, green), each involving twelve nodes; the corresponding connectivity matrices are shown in the two insets. The orange and yellow lines in the background of the main figure are the hyperbolic and elliptic Lagrangian structures for the twelve-hour time window under consideration, which are ultimately responsible for the discrepancy in the two different connectivity matrices.

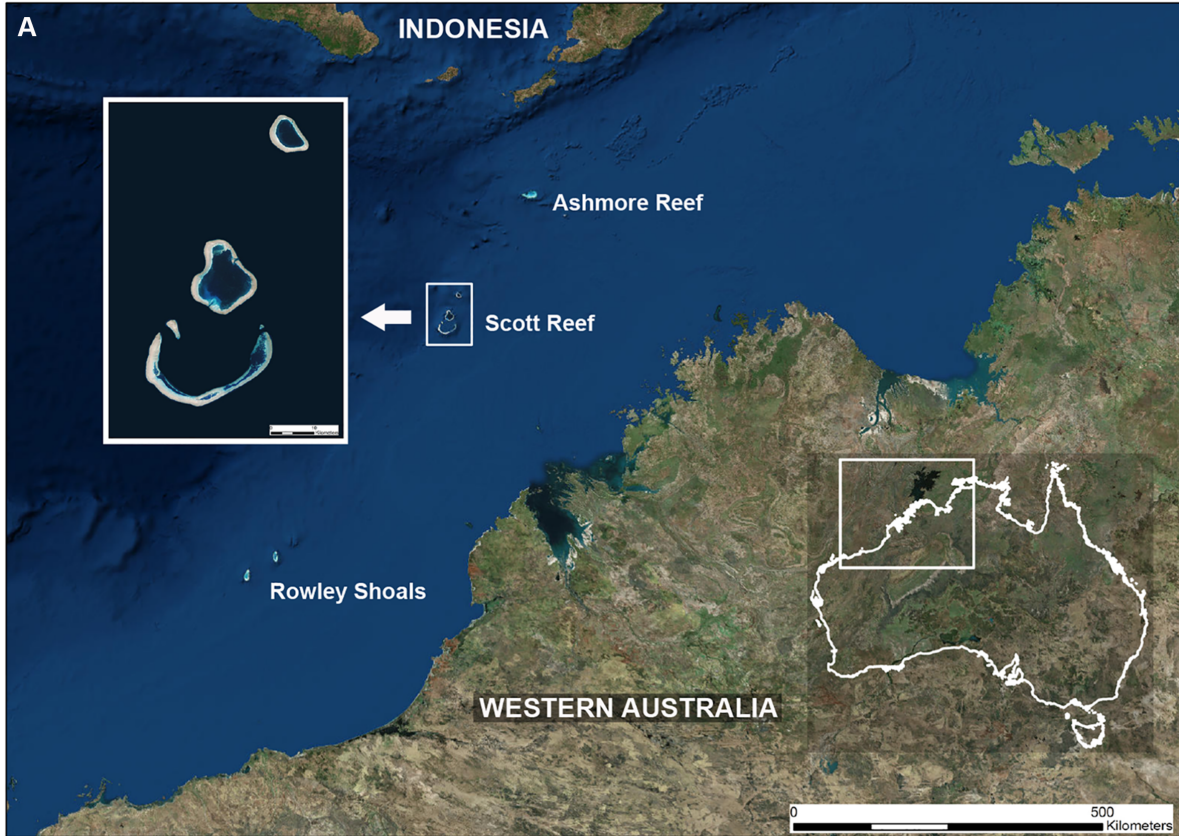


Figure 3-3: Location of the Scott Reef atoll system off the northwest coast of Western Australia (maps accessed in ESRI, ArcGIS). Figure and caption from Foster & Gilmour [2018].

## 3.2 Scott Reef geography

Scott Reef is a remote coral atoll system in Western Australia, roughly half-way between the northwestern coast of Australia and Indonesia. A satellite image and its location on a map of Australia are shown in figure 3-3 on page 96. Scott Reef emerges from depths greater than 2,000m on the edge of the continental shelf, as shown in figure 3-4 on page 97. It is comprised of two atolls: Scott Reef North, an almost-closed lagoon, and Scott Reef South, a crescent-shaped atoll with a shallow half-lagoon with a 12-km diameter. The particular study area is the channel between Scott Reef South and Sandy Islet, an emerged patch of sand. The channel, emphasized in figure 3-4 by the red square, is about 2.8 km in width and 4.3 km in length.

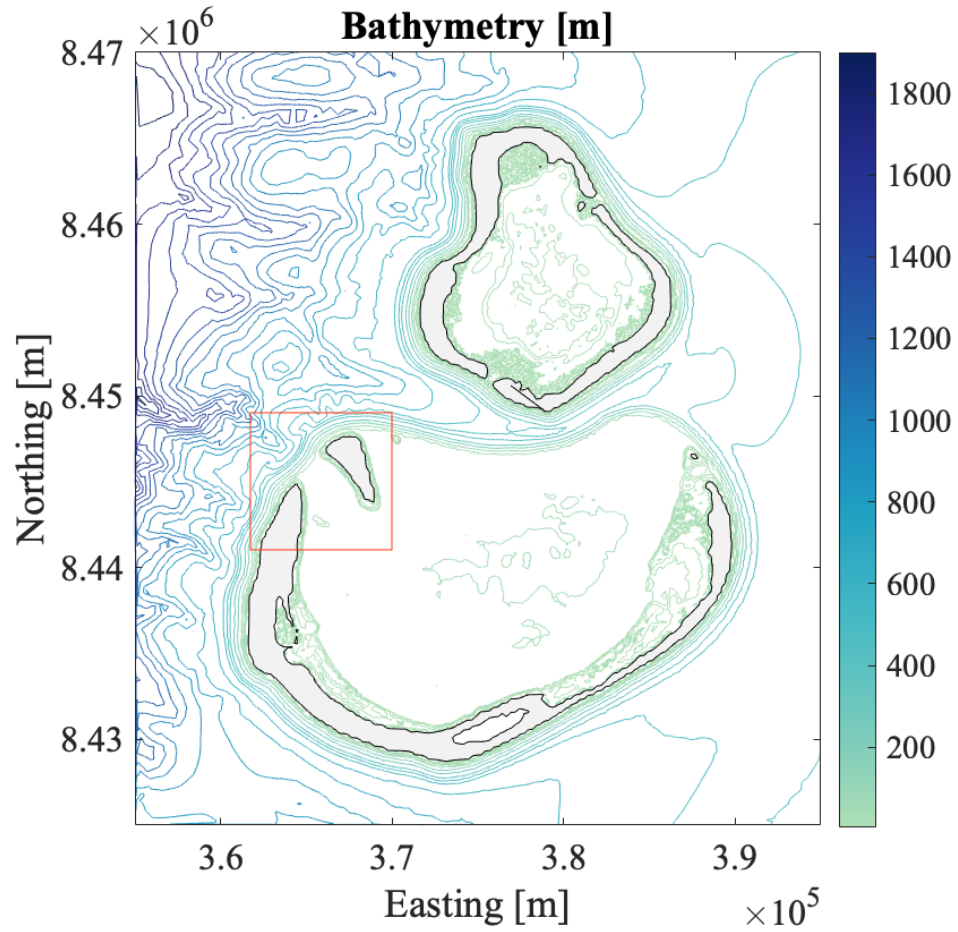


Figure 3-4: Bathymetry of Scott Reef from Rayson *et al.* [2018]. The red square indicates the channel between Scott Reef South and Sandy Islet.

The flow at Scott Reef South is macrotidal, with ranges up to 4 m, and mainly driven by the forcing from barotropic tides flowing through the channel [Rayson *et al.*, 2018]. This tidally-driven channel flow controls the heat exchange between the offshore waters and the lagoon waters, as well as its supply of nutrients [Green *et al.*, 2019]. The lagoon is delimited by the bathymetric contours in figure 3-4. The channel is also characterized by the steep bathymetric gradient, with depths going from 10 m to 1,000 m within 5 km. Scott Reef’s remote location has protected the system from many direct anthropogenic threats. As a biodiversity hot spot, Scott Reef is frequently monitored and extensively studied, with numerical models augmented by regular hydrographic surveys [Rayson *et al.*; Gilmour *et al.*, 2013; Green *et al.*, 2019].

### 3.3 Numerical modeling of Scott Reef

The numerical ocean model of the region used for the study was produced by Rayson [2011]. The model was based on SUNTANS, a three-dimensional, non-hydrostatic circulation model that uses Reynolds averages and the Boussinesq approximation on an unstructured grid. The horizontal resolution around Scott Reef was 75 m and the temporal resolution was 30 minutes. Details on the model can be found in Rayson *et al.* [2018]. The barotropic tide solution was derived from the TPXOv7.2 global model. The boundary conditions were specified as tidal fluxes along the boundary edges or by specifying the free-surface directly in the boundary cell centers, with the tidal fluxes interpolated from the solutions of two models, HYCOM and BRAN. HYCOM is the HYbrid Coordinate Ocean Model, which is part of the Global Ocean Data Assimilation Experiment (GODAE) that is sponsored by the National Ocean Partnership Program (NOPP). BRAN stands for Bluelink ReANalysis, a velocity output of the data-assimilating Ocean Forecasting Australia Model (OFAM). The domain of the numerical model spanned almost 600 km of the Kimberly Shelf region, with Scott Reef located roughly at the center of this domain. The area of interest was the channel



northwest of Scott Reef South.

The planning of the field experiment relied on the Lagrangian analysis of the model velocity field data. This predictive analysis was carried on a dataset output by a previously existing model run from Rayson [2011]. The run was narrowed between September 27 and October 30, 2007 to output surface velocity conditions at the same time of the year that the cruise was planned to take place, around October 2016.

Following the cruise, a more recent SUNTANS run was output with parameters corresponding to the sea conditions experienced during the cruise, which took place between September 29, 2016 and October 16th, 2016. Unless otherwise noted, all figures shown here are based on the 2016 model run, which benefitted from an improved mesh and data assimilation from additional hydrographic surveys. The meshes used for the two iterations of the model are shown in figure 3-5 on page 100. The most obvious change is the addition of the coastline of Sandy Islet. The bathymetry around the channel was also refined. The SUNTANS model did not incorporate any wind forcing as it uses a rigid-lid approximation; the observed windage during the cruise was very low.

All model runs were generated by Matthew Rayson and the resulting velocity fields were output in netCFD format. For the predictive analyses using the 2007 dataset, which are described in the following subsection, the FTLE fields were computed by Alireza Hadjighasem using the method by Farazmand & Haller [2012]. For the preliminary analyses of the 2016 dataset, I computed the numerical trajectories and FTLE fields following the same method. For ease and speed of calculations, the model data was subsequently uploaded in netCFD format to a web-based gateway by Ameli & Shadden [2019], TRACE, which stands for Trajectory Reconstruction and Analysis for Coherent Structure Evaluation. TRACE uses parallel processing for the Lagrangian computations and can handle and process very large datasets that are hosted remotely. More information can be found at <http://transport.me.berkeley.edu/restore>. TRACE can output the advection of numerical trajectories and the corresponding FTLE fields. For the results presented in this chapter corresponding to

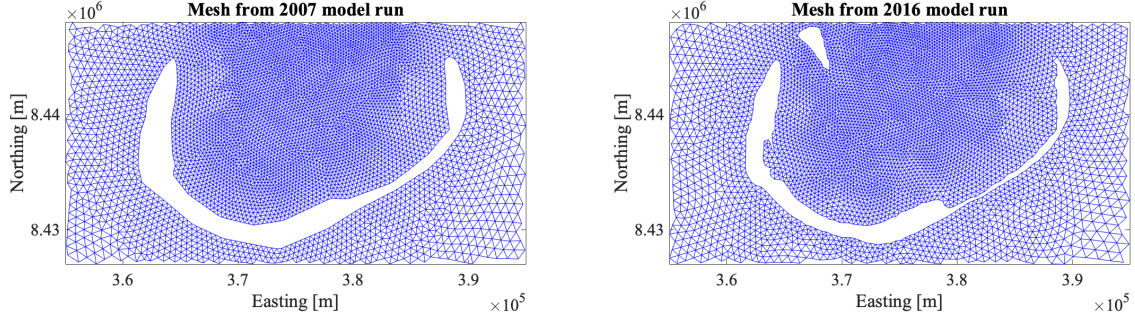


Figure 3-5: Meshes used for the SUNTANS model runs.

the 2016 dataset, the numerical trajectories were all generated in TRACE.

### 3.4 Predictive analysis of the 2007 dataset

To get a first understanding of the flow organization in the area during the time period of interest, the FTLE fields were computed by Alireza Hadjighasem over the time domain at one-hour increments. The purpose of the predictive FTLE analysis was to look at high rates of particle separation in the fluid flow domain and identify any key LCS structure governing the flow. Forward-time FTLE processing of the 2007 data revealed the formation of a pronounced and recurring pattern forming in the channel between Scott Reef South and Sandy Island around high tide, an example of which is shown for September 27, 2007 in figure 3-6 on page 101. The pattern of interest is the formation of continuous curves of high FTLE values, commonly referred to as FTLE ridges. The ridges, here, are extracted by thresholding on high values of the normalized FTLE field and they indicate the locally highest rates of separation. Such ridges were detected around the eastern tip of Scott Reef South as well, but these features lacked the periodicity of the ridges in the channel, which were recurring at high tide. FTLE ridges were formed throughout the ebb, going from high tide to low tide, first being formed southeast of the channel, then progressively northward with the ebb. The FTLE features were the most pronounced with an integration time-

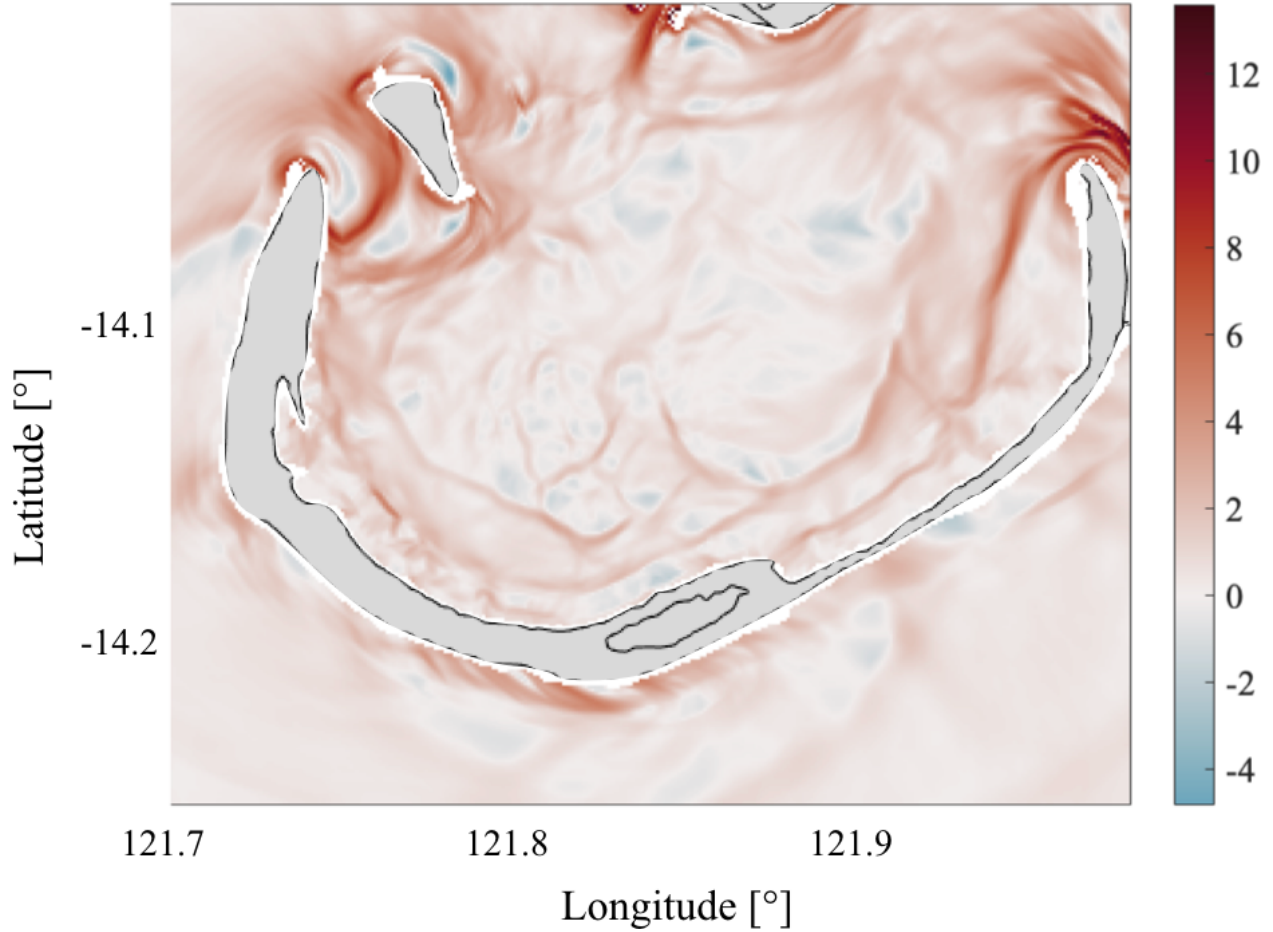


Figure 3-6: FTLE analysis of the 2007 dataset for Scott Reef South. The FTLE field is shown for September 27, 2007 for the period from 00:00:05 to 06:00:05 UTC.

interval of 6 hours, a time scale that is physically natural and relevant for a system governed by the semi-diurnal tide cycle. On top of exhibiting interesting FTLE features, the channel is of high importance for physical and biological oceanographers, as previously mentioned [Green *et al.*, 2019; Foster & Gilmour, 2018], thus it became the focus area of this study.

Prior to the field experiment, the complimentary spectral clustering analysis was carried out in the conventional way by Alireza Hadjighasem focusing on the channel, the targeted area for the Lagrangian study with a domain of  $0.04^\circ$  by  $0.04^\circ$ , as shown in figure 3-7 on page 102. The analysis was conducted for a 6-hour integration window to match the FTLE computations. The spectral clustering results, shown in figure 3-7.a, were two clusters, one northwest (yellow) and one southeast (turquoise), separated by a background cluster cor-

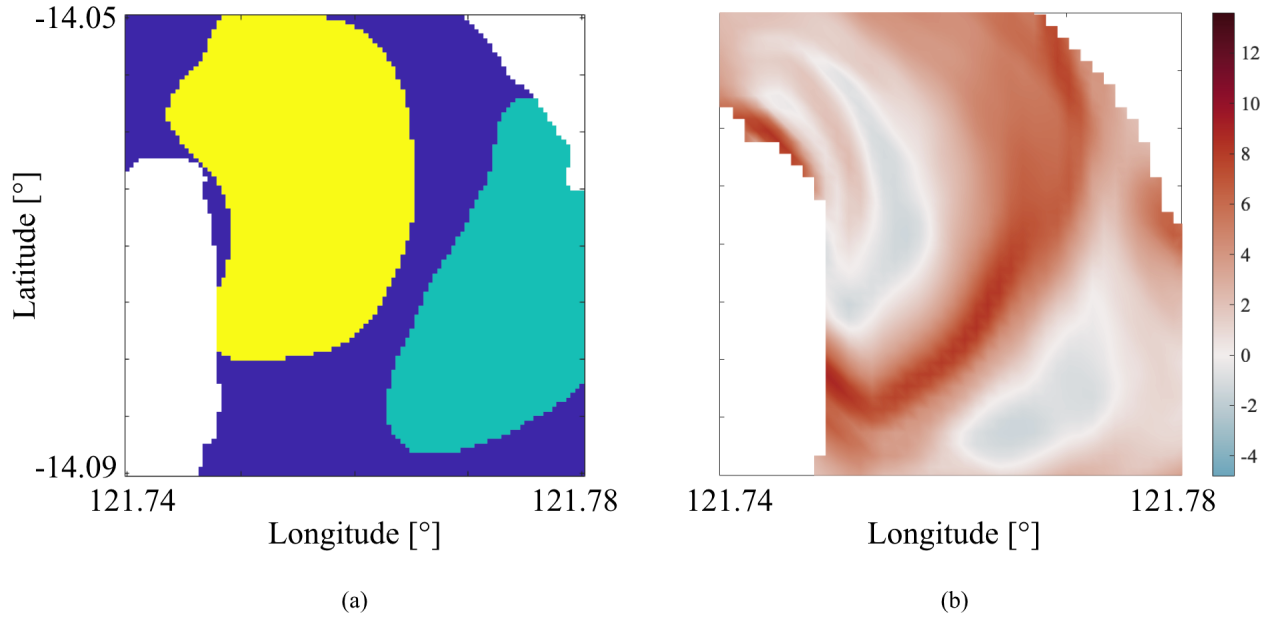


Figure 3-7: (a) Spectral clustering and (b) FTLE analyses of the 2007 dataset for the channel. Data from Alireza Hadjighasem.

responding to the incoherent region (dark blue), as shown in figure 3-7.b. The clustering results, including the size and number of clusters, were computed with the conventional spectral clustering method from Hadjighasem *et al.* [2016] and chosen without conducting a sweep of parameters. The cluster coherence metric was not available when this pre-experiment model analysis was conducted.

The preliminary spectral clustering analysis suggested that the flow domain was partitioned into lagoon waters, southeast of the FTLE ridge, shown with the turquoise cluster, and offshore waters, northwest of the ridge, shown with the yellow cluster. Based on these predictions, the expectation was that drifters released in the vicinity of the channel would separate into three batches according to their initial release location relative to the ridge: a northwestern batch and a southeastern batch of drifters, with drifters released within the mixing region having a different transport behavior. Another expectation was that for the 6-hour duration of the experiment, drifters released on either side of the ridge would stay either within the lagoon waters or within the offshore waters.

For the FTLE ridges reoccurring around high tide, an ideal experiment would encompass

a drifter release at high tide around spring tide and at high tide around neap tide. Due to constraints on the ship's schedule, however, the research vessel was in the vicinity of the channel for about 5 days only, around Spring tide. The drifter experiment thus focused on a high tide event around Spring tide.

### 3.5 Field experiments

The drifter experiment was conducted as part of a cruise organized by the Australian Institute of Marine Science aboard research vessel (R/V) Solander, shown in picture 3-8 on page 104. The cruise departed Broome, Western Australia on September 29, 2016 and returned to port on October 16th, 2016. R/V Solander was in the vicinity of the Scott Reef channel from October 1, 2016 to October 6, 2016. For this experiment, I built the drifters in collaboration with Carlin Bowyer at the University of Western Australia. The drifters were then shipped to Broome and I assembled them aboard the ship before the release. A total of 24 drifters were built for the experiment.

The drifters consisted of two parts: a surface drogue and a submerged sea anchor. A picture of a drifter and its component is shown and annotated in figure 4-3 on page 138. This in-house, custom-made design is not as Lagrangian as the standard CODE/DAVIS drifters from Davis [1985], but the CODE drifters were not available as they are much more expensive and much harder to carry on a medium-sized vessel. The surface drogue was a cylindrical pipe allowing a commercially available GPS tracker, the SPOT Tracker model, to emit position and time signals. The buoyancy of the surface drogues was adjusted to leave only about 20 cm emerged, so that the windage on the drifter would be minimized. The off-the-shelves sea anchor was immersed at about 2-meter deep to minimize surface drag. The SPOT trackers were set to emit at 2-minute intervals. The trackers - and thus drifters - positions over time were retrieved through the SPOT online portal once the ship returned to shore, as Internet connection was limited aboard the ship.



Figure 3-8: Research Vessel Solander from the Australian Institute of Marine Science docked in Broome, Western Australia.



Figure 3-9: Surface drifters built *ad hoc*

To demonstrate the occurrence of the FTLE ridge and the spectral clusters, arrays of surface drifters were released throughout the channel around high tide during Spring tide. For experimental repeatability and to allow a pilot experiment, the experiment was split into two releases, each with two arrays of drifters. Because the spring tide in October 2016 occurred at night and the ship did not allow nocturnal operations, the first release happened at the diurnal high tide before spring tide and the second release at the diurnal high tide after spring tide. Before the first release, out of 24 drifters and GPS trackers on board, 3 trackers failed to communicate with the satellite. The remaining 21 functional trackers were split into two deployments consisting of 9 and 12 trackers.

For each deployment, two 8-kilometer-long parallel arrays were released roughly orthogonal to the forward FTLE ridge, around its predicted time and location. The release locations are shown in figure 3-10 on page 106. The drifters were released from North to South to capture the separation feature at all times, as an FTLE ridge was predicted to be formed progressively more northwest over the ebb: drifters released north of the ridge would capture the northern spectral cluster before it was advected northward, drifters released in the middle of the array would fall on either side of the forward FTLE ridge while drifters released south of the ridge would remain south. The drifters were released in a zig-zag pattern to minimize path length and thus ship time: in spite of this effort, the total deployment nevertheless lasted up to 1.5 hour, out of the 6-hour-long ebb.

## 3.6 Results

An overview of the key experimental results is provided in section 3.6.1. The FTLE analysis and the spectral clustering analysis are presented in sections 3.6.2 and 3.6.3, respectively. Additional details on the FTLE ridges computations are provided in appendix B.



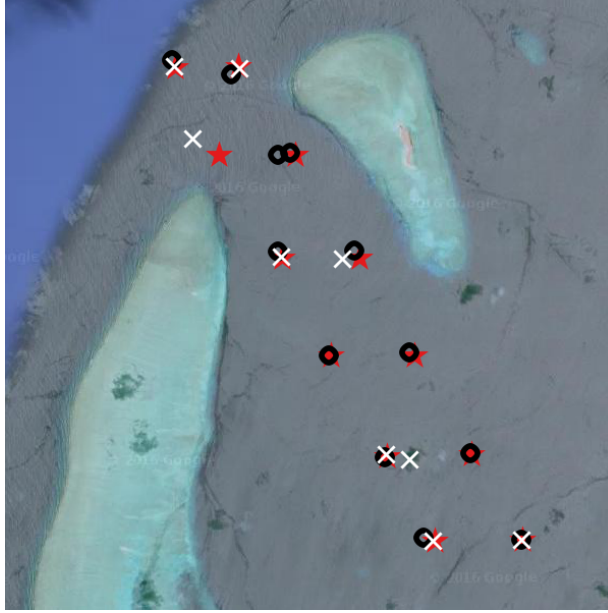


Figure 3-10: Release strategy for the drifters at Scott Reef: (red stars) planned initial positions, first satellite signals for the first (white crosses) and second (black circles) releases.

### 3.6.1 Summary of the key experimental results

For the primary drifter release, the first drifter's position was retrieved at 06:15:00 UTC, about 15 minutes after the start of the ebb. For the last drifter in water, the initial position was at 07:30:00 UTC. The main results of the release are presented in figure 3-11 on page 107, which plots the drifter trajectories from their initial positions until 13:30, 6 hours after the last drifter was deployed. Although the SPOT trackers were set to emit at a two-minute frequency, the time intervals of received position were very irregular, so all plots show positions within 5-minute intervals of the corresponding times.

The key result from figure 3-11 is that the drifters were separated into two groups: a northern batch, marked in blue, that was expelled offshore the channel and a southern batch, marked in orange, that remained within the lagoon waters. This partition of trajectories into two batches validated the FTLE predictions that the domain would separate into a northern and southern subdomain. The spectral clustering analysis predicted that the drifters would separate into three clusters, however. The initial drifter positions are indicated with bold markers and their final positions are indicated by their number, corresponding to the order



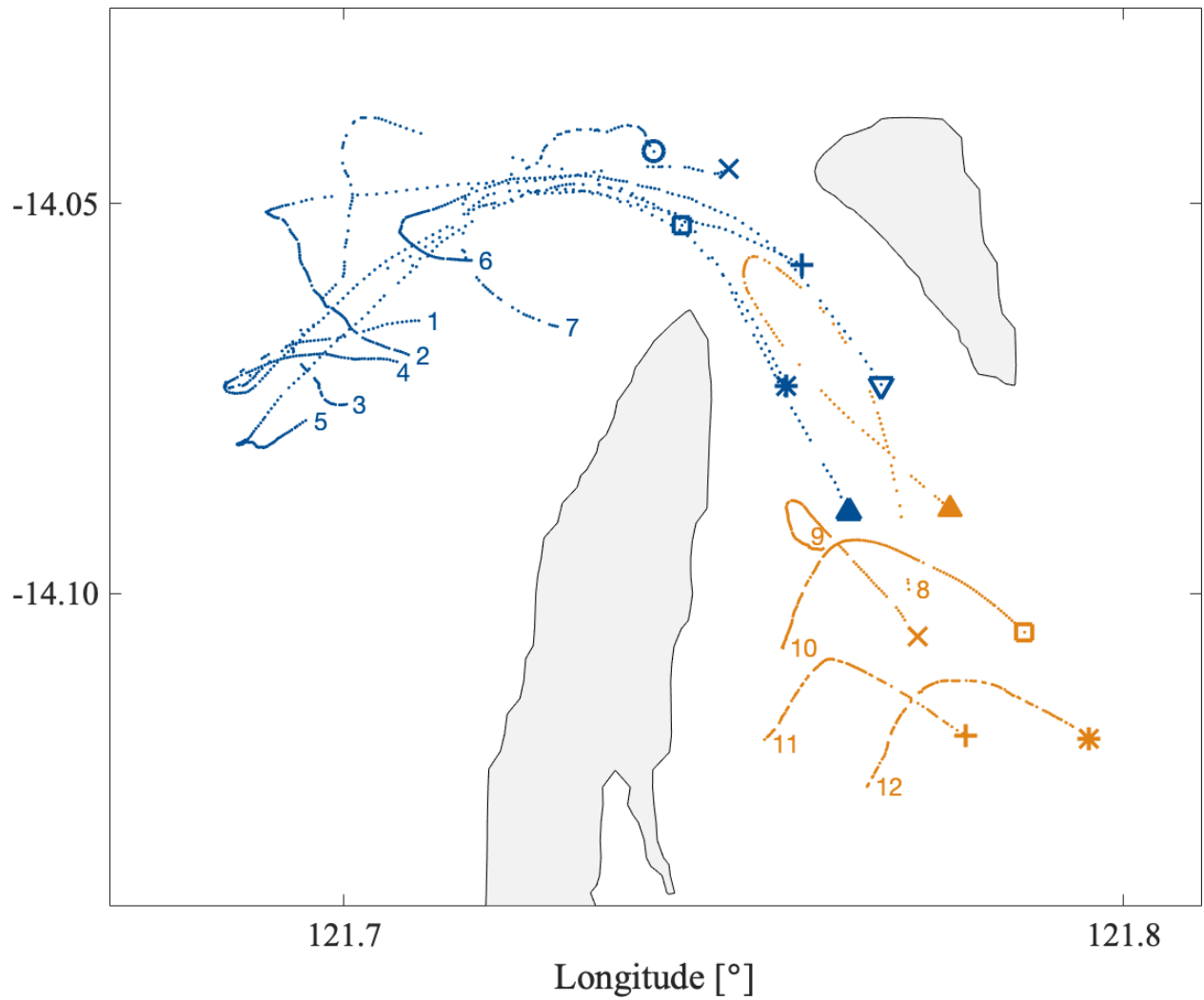


Figure 3-11: Scott Reef field release: drifter trajectories. The initial drifter positions are indicated by the bold markers. The final positions are indicated by the numbers, corresponding to the order of release. The trajectories of drifters that ended offshore after 6 hours are colored in blue, whereas those of drifters that stayed within the lagoon are colored in orange.

of deployment. The order of trajectories shows the zig-zag pattern from north to south. Drifters 7 and 8, indicated by filled triangles, were released in close proximity but their trajectories had very different behavior: they both initially started leaving the channel, but drifter 8 looped around to stay within the lagoon whereas drifter 7 ended offshore.

The time-evolution of the Eulerian velocity field and the LCS structures recalculated from the 2016 model data are presented in figures 3-12–3-13 on pages 110–111, along with the drifter positions throughout the experiment. The 6-hour FTLE fields were computed in forward- and backward-time on sliding time windows from 06:00:00 to 12:00:00 UTC. The coherent cluster inside the channel was advected from 06:00:00 to 07:00:00. The spectral results were recalculated at 07:30:00, time at which all drifters has been deployed. The velocity fields are shown in panel (a); the drifter positions, the computed forward-time FTLE ridges results and the spectral clusters are shown in panel (b); and the computed backward-time FTLE ridges results, spectral clusters from (b) and the numerical trajectories are shown in panel (c). The numerical trajectories were seeded individually to match the initial conditions of their corresponding drifter. The computations for the FTLE ridges are explained in appendix B, along with the comparison between the advected ridges and the ridges calculated on a sliding integration window.

In figure 3-12, the times are shown around the initial deployment time at 06:30:00 UTC and half-way through the deployment at 07:00:00. The spectral clusters were advected from 06:00:00 and the FTLE ridge was recomputed at each time. The FTLE analysis is described in section 3.6.2 and the comparison between advected and recomputed ridges is discussed in appendix B. The spectral clustering results from 06:00:00 are discussed in section 3.6.3. In figure 3-13, the results are shown at the beginning to the release at 07:30:00 UTC, once all drifters were in water; around the turn of the tide at 10:00:00 UTC, which is roughly half-way through the 6-hour experiment; as well as at low tide at 12:00:00 UTC and, finally, at 13:30:00 UTC, 6 hours after the last drifter was in water. The clusters are advected from 07:30:00 and the ridges are recomputed at every time. The advection of FTLE ridges

generating a lot of noise, the recomputed ridges are plotted from 10:00:00 to 13:30:00; for this reason, they are displayed in grey. By 07:30:00, a second cluster was formed south of the ridge and inside the lagoon, indicating that drifters released south of the ridge would stay within the lagoon.

Figures 3-12–3-13 illustrate how the drifters progressively split into two batches, one north (blue) and one south (orange): the first 7 drifters (in blue markers) were released north of the forward FTLE ridge (red curve) and the last 5 drifters (orange markers) were released south of the ridge. The first 6 drifters were released within the blue cluster. The seventh (filled blue triangle) and the eighth (filled orange triangle) were released in close proximity, but on different side of the ridge. Drifters 8 through 12 were released inside the orange cluster that had just formed inside the lagoon. Drifters 8 and 9 stayed in close proximity until 10:00:00, around the turn of the tide, time at which the eighth drifter looped around towards the lagoon: indeed, it was located south of the repelling FTLE ridge when it was released. At 10:00:00, the FTLE ridge is indicated by the gray line, as it is the ridge corresponding to the 10:00:00 to 16:00:00 time window. Drifter 8 however was always positioned south of this ridge. The blue batch of drifters, which was located north of the evolving FTLE ridge, got expelled out of the channel, while the orange batch stayed within the lagoon, as predicted from the forward FTLE fields. The first eight drifters were released close to very high backward FTLE values, meaning high rates of attraction, indicated by the ridges plotted in green. These drifter positions closely followed the backward ridge. These trajectories deviated from the ridge after the turn of the tide at 10:00:00, however, and were carried past the FTLE ridge, likely carried by inertia.

The discrepancy between the experimental drifter trajectories and the evolution of the clusters is likely due to the differences between the numerical model and the real oceanic flow. While the drogues were floating at the surface, the anchors were submerged at 2 meters, whereas the surface velocities were output by the model for the zero-meter isobar. The discrepancies were especially pronounced for the northern, blue group of drifters: the

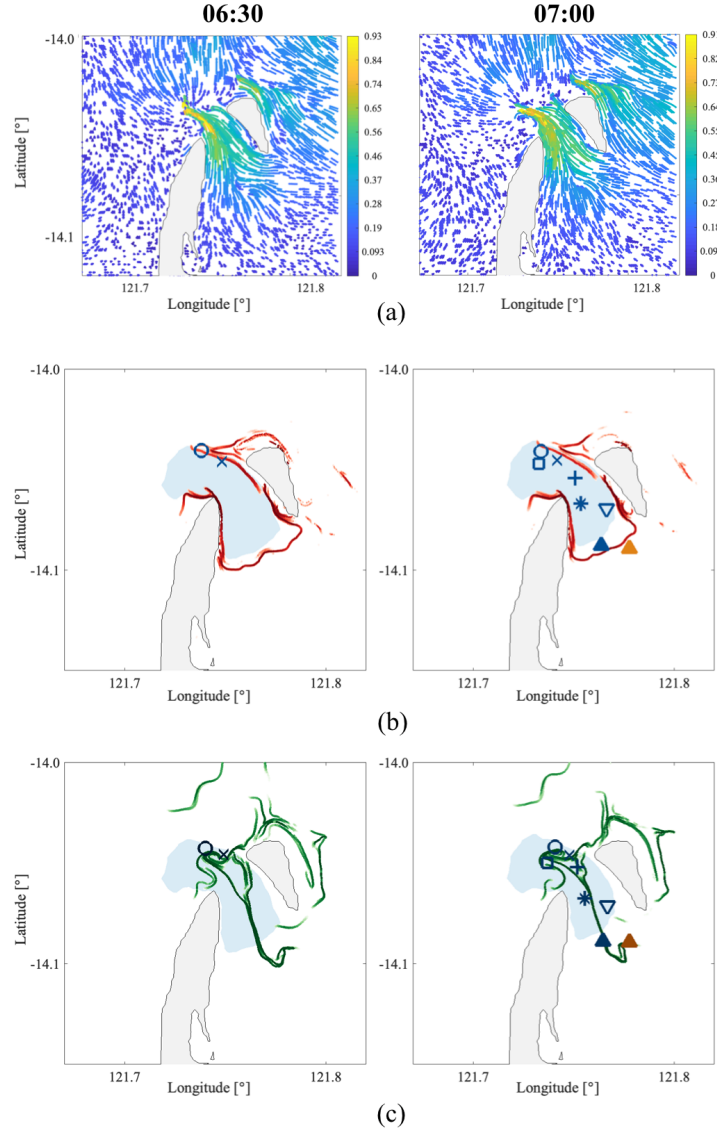


Figure 3-12: Scott Reef field release Part 1. Comparison between the experimental drifter positions, the velocity field and the LCS analysis throughout the release at (Left) 06:30, (Center) 07:00 and (Right) 07:30, time at which all drifters were in water. (a) Eulerian field of velocities in m/s. (b) Spectral clusters, forward FTLE ridge and experimental drifters positions. The clusters from 06:00 were advected to 06:30 and 07:00. (c) Spectral clusters from (b), backward FTLE ridge and numerical drifter positions.

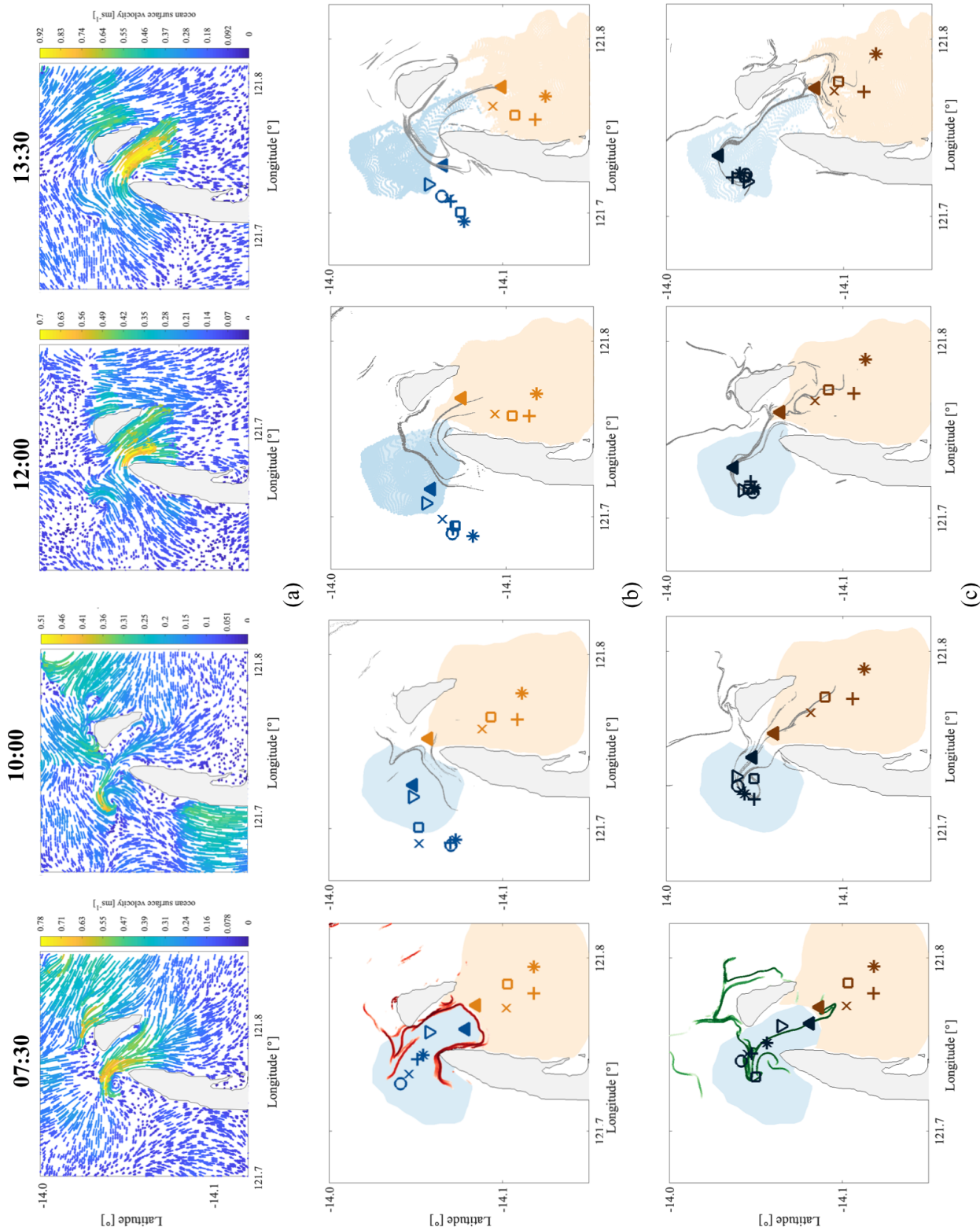


Figure 3-13: Scott Reef field release Part 2. Comparison between the experimental drifter positions, the velocity field and the LCS analysis throughout the release at (from left to right) 07:30, time at which all drifters were in water; 10:00, roughly half-way through the 6-hour experiment; 12:00 at low tide and start of the flood; and 13:30, 6 hours after all drifters were in water. (a) Eulerian field of velocities in m/s. (b) Spectral clusters, forward FTLE ridge and experimental drifters positions. The clusters were recomputed at 07:30 and advected from there. (c) Spectral clusters from (b), backward FTLE ridge and numerical drifter positions.

cluster was smaller than the lagoon cluster, leaving more room for deviations in trajectories, and the velocities were much higher offshore than within the lagoon. The southern cluster comprising almost all of the lagoon waters, it was expected that the southern batch of drifters stayed within. The northern group of drifters was carried past the location of the blue cluster by 12:00 and 13:30. Another probable cause for the *overshooting* of the drifters is the momentum that they carried, especially in such channel flows where there is a strong reversal of the tide. In both the model and in the real world, however, the expelled drifters stayed somewhat grouped. To further investigate the differences between model and real world flows, numerical trajectories were computed. They are plotted superimposed to the backward FTLE ridge in panel (c) of figures 3-12–3-13. All numerical tracers that were released within the blue spectral cluster stayed within the cluster, as expected. The numerical particles ended up more dispersed at 13:00:00 than the northern batch. Neither the FTLE features nor the presence of coherent spectral clusters cannot be discerned from the Eulerian velocity fields, particularly for the forward-time FTLE ridge. Details of the 2016 Lagrangian analysis are explained in the following subsections.

### **3.6.2 FTLE analysis of the 2016 dataset: neap tide versus spring tide events**

The FTLE analysis presented in figures 3-12–3-13 and below in this section was conducted on the updated 2016 model run. Both forward and backward time analyses were conducted to look at the repelling and attracting structures within the flow. The analysis was first conducted over the entire domain of Scott Reef South, as illustrated in figure 3-14 on page 113, for both neap and spring tide events. For the time period of interest, around the time of the R/V Solander cruise, neap tide occurred on September 27, 2016 and high tide occurred on October 3, 2016. At a critical point of the tidal cycle, at high tide during spring tide, a pronounced forward-time, repelling FTLE ridge was formed south of the channel. The backward-time FTLE computations revealed the presence of an attracting ridge within the

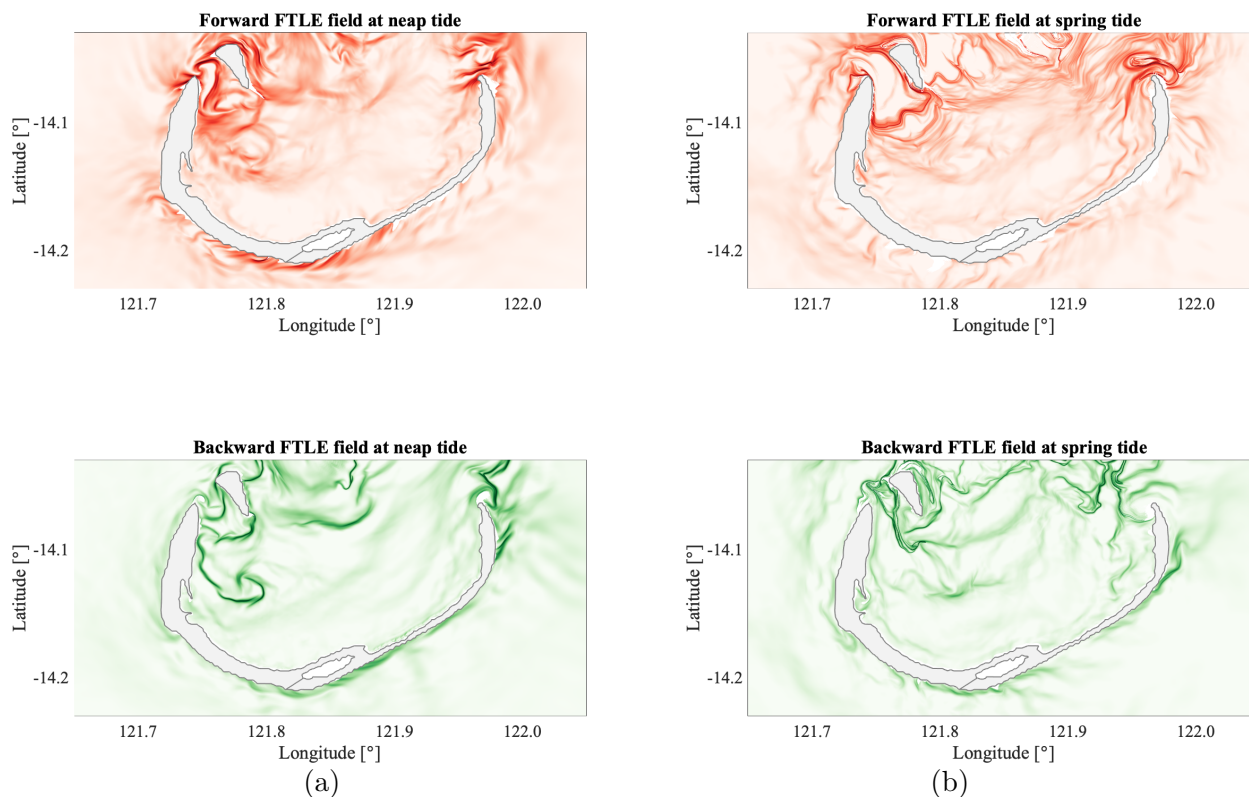


Figure 3-14: FTLE fields in forward and backward times for Scott Reef South around (a) neap tide on September 27 and (b) spring tide on October 3. The period of integration was 6 hours in all cases.

channel. The highest values of backward-time FTLE were found in the channel at both neap and spring tides. Similarly to the 2007 analysis, ridges were present elsewhere within Scott Reef South, but the ridges within the channel were more pronounced and occurred in a recurring manner throughout the Spring tide ebb, from high to low. The formation of sequential ridges formed progressively northward over sliding windows of integration within the channel is shown in figures 3-12–3-13. At high tide during neap tide, the features within the channel were not as sharply defined.

Focusing the analysis on the channel between West Hook and Sandy islet, figure 3-15 on page 115 shows the FTLE fields in forward and backward time at high tide during ebb and neap tide within the zone of interest. The FTLE values were much higher at spring tide than at neap tide, with the ridges being more sharply defined. The negative FTLE values indicate

significant vertical motion, which will be investigated in a subsequent section. These negative values of FTLE were also much stronger during spring tide than during neap tide. For all subsequent analyses, this domain of  $0.15^\circ$  by  $0.15^\circ$  around the channel was used. Around spring tide, the FTLE field in forward time reveals the formation of a pronounced FTLE ridge where rates of separation are maximum between the lagoon waters and the offshore waters. In backward time, the FTLE field indicates strong attraction along the channel. Note that here, the FTLE ridges refer to the local maxima; they are self-intersecting and thus do not correspond to manifolds. The ridges are extracted for clarity of visualization only. The FTLE fields predict that during the ebb, going from high tide to low tide, the lagoon waters stay trapped within the lagoon, whereas the water mass within the channel and north of the forward FTLE ridge gets expelled outside the lagoon following the backward FTLE ridge.

The formation of forward-time FTLE ridges occurring sequentially during the ebb around spring tide was observed for 4 consecutive days within the numerical model. The recurrence of this ridge in the channel is shown in figure 3-16 on page 116. Starting at high tide around spring tide, as the flow starts leaving the Scott Reef South lagoon through the channel, a ridge is formed at the southeastern part of the channel, almost normal to it. As the ebb progresses, over sliding windows of integration, a ridge is formed progressively more north and west along the channel. At the reversal of the tide at low tide, a ridge is formed northwest of the channel, as shown in figure 3-16. The FTLE ridges were strongest at spring tide, particularly at high tide around spring tide, about two days prior to and two days after the spring tide event. The neap tide event did not exhibit the repeatability of the spring tide event: the ridges did not occur periodically as strongly and sharply as they did for the spring tide event.



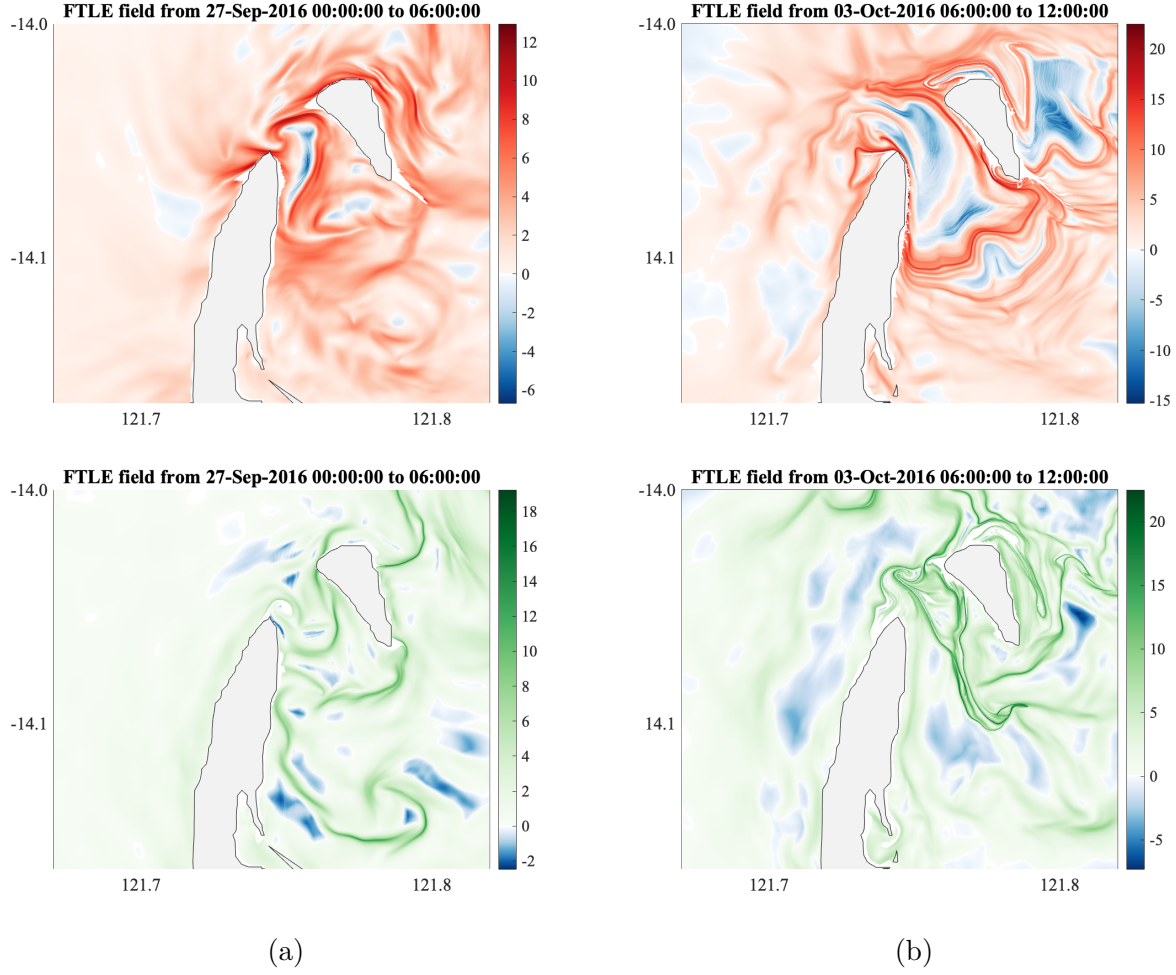


Figure 3-15: FTLE fields in forward and backward times in the Scott Reef channel around (a) neap tide on September 27 and (b) spring tide on October 3. The period of integration was 6 hours in all cases. In forward time, high FTLE values (dark red) indicate high rates of separation between particles. In backward time, high FTLE values (dark green) indicate high rates of attraction.

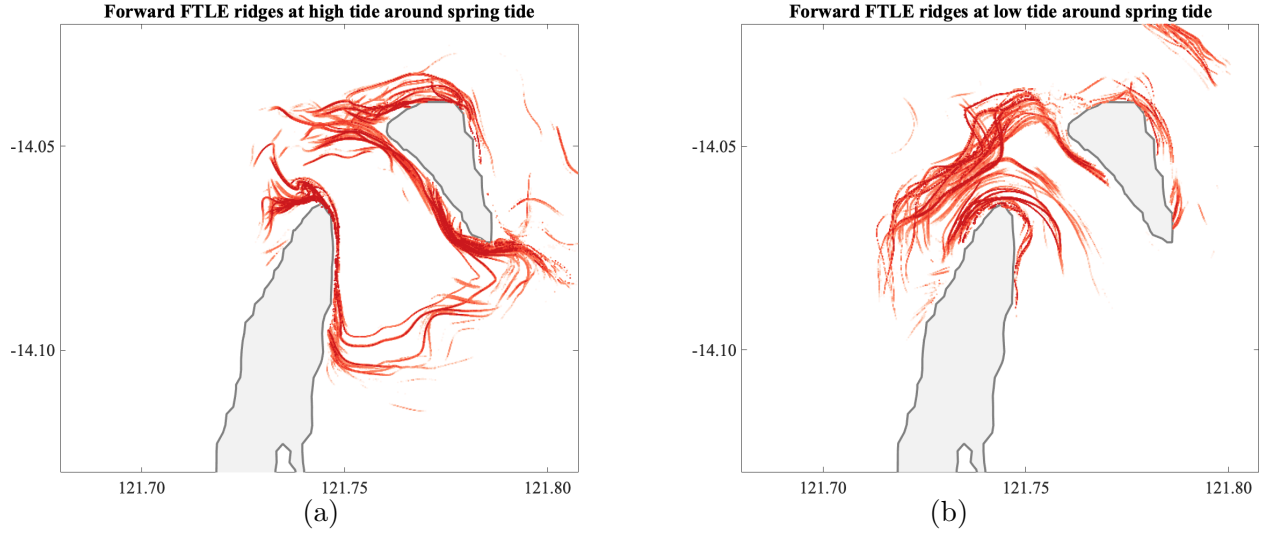


Figure 3-16: Periodicity of the FTLE ridges around spring tide: superimposition of FTLE ridges (dark red) at 6 consecutive (a) highs and (b) lows of the tidal cycle.

### 3.6.3 Spectral clustering analysis of the 2016 dataset: finding the optimal parameters

Having detected the areas of high separation rates, the parameter-free spectral clustering analysis was carried out as a complimentary method to search for coherent clusters within the domain. The protocol described in chapter 2 was followed to analyze the neap and spring tide events. The spatial and temporal domains were the same as for the FTLE, but the resolution was much lower for computational purposes, with a resolution of  $6 \times 10^{-4}^\circ$ . The period of integration of particle trajectories was kept at 6 hours.

Following the first step of the spectral clustering protocol, the sweeps of  $r$  radii are in shown in figure 3-17 on page 118. For the neap tide event, a peak in gap ratio was detected at  $r = 0.0250$ , as shown in figure 3-17.a. For the spring tide event, a peak in gap ratio was detected with for  $r = 0.0175$ , shown in 3-17. The spectral clustering results were calculated for the 6-hour time window starting at 07:30, time at which all drifters were in water. The ebb lasting 6 hours, a 90-minute drifter deployment meant that the flow was quite different at the start of the 6-hour experiment than from high tide. The sweep of  $r$  parameters showed a peak at 0.0325. The gap ratios for the 06:00 flow were much higher on average than for

the 07:30 flow or for the neap tide event, with the peak gap ratio being almost twice as high as the other peaks. In all cases, convergence in the spectral shape, step 2 of the protocol, was obtained after an offset coefficient of  $10^7$  for all peaks, as illustrated in figure 3-18 on page 119.

The clustering results and the noise coherence metrics were calculated for all peak values. The results are shown in figure 3-19. Although the values of  $r = 0.025$  and  $r = 0.0175$  were comparable in magnitude, the respective clustering results for the neap and spring tide events were very different. At neap tide, the  $r$  value of 0.025, yielded 18 adjacent clusters, shown in figure 3-19.a. This partition of the domain indicates that there was no component within the domain that was particularly strongly connected, i.e., the domain was rather homogeneous in connectivity. The incoherent background cluster is impossible to extract from this picture alone. All coherence percentages fell within 94 and 100%. At high tide around spring tide, the gap ratios were on average much higher than in the neap tide case. The  $r$  value of 0.0175 yielded 2 coherent clusters very distinct from the incoherent background cluster. The two clusters had a coherence metrics of 91.39% for the cluster east of Sandy Island and of 94.96% for the cluster inside the channel. After 6 hours, both clusters significantly decreased in size, which will be investigated in the next section. At the 07:30 start of the release, five coherent clusters were detected: two clusters roughly corresponding to the two clusters from 06:00, one lagoon cluster and two offshore. The cluster in the northern part of the channel roughly corresponded to the cluster from 06:00, although it was larger at 07:30, encompassing a bit of the offshore domain. Its coherence percentage was within 3% of that of the 06:00 cluster. The main difference is that it experienced no noticeable shrinkage. Similarly, the cluster east of Sandy Islet corresponded to the second coherent cluster from 06:00. The lagoon cluster was large enough to englobe almost all of the lagoon waters within the domain and had a coherence value of 92.76%. It also experienced minimal shrinkage. The two offshore clusters had coherence values of 64.8 and 84.12%, much lower than the clusters within the channel, the zone of interest.

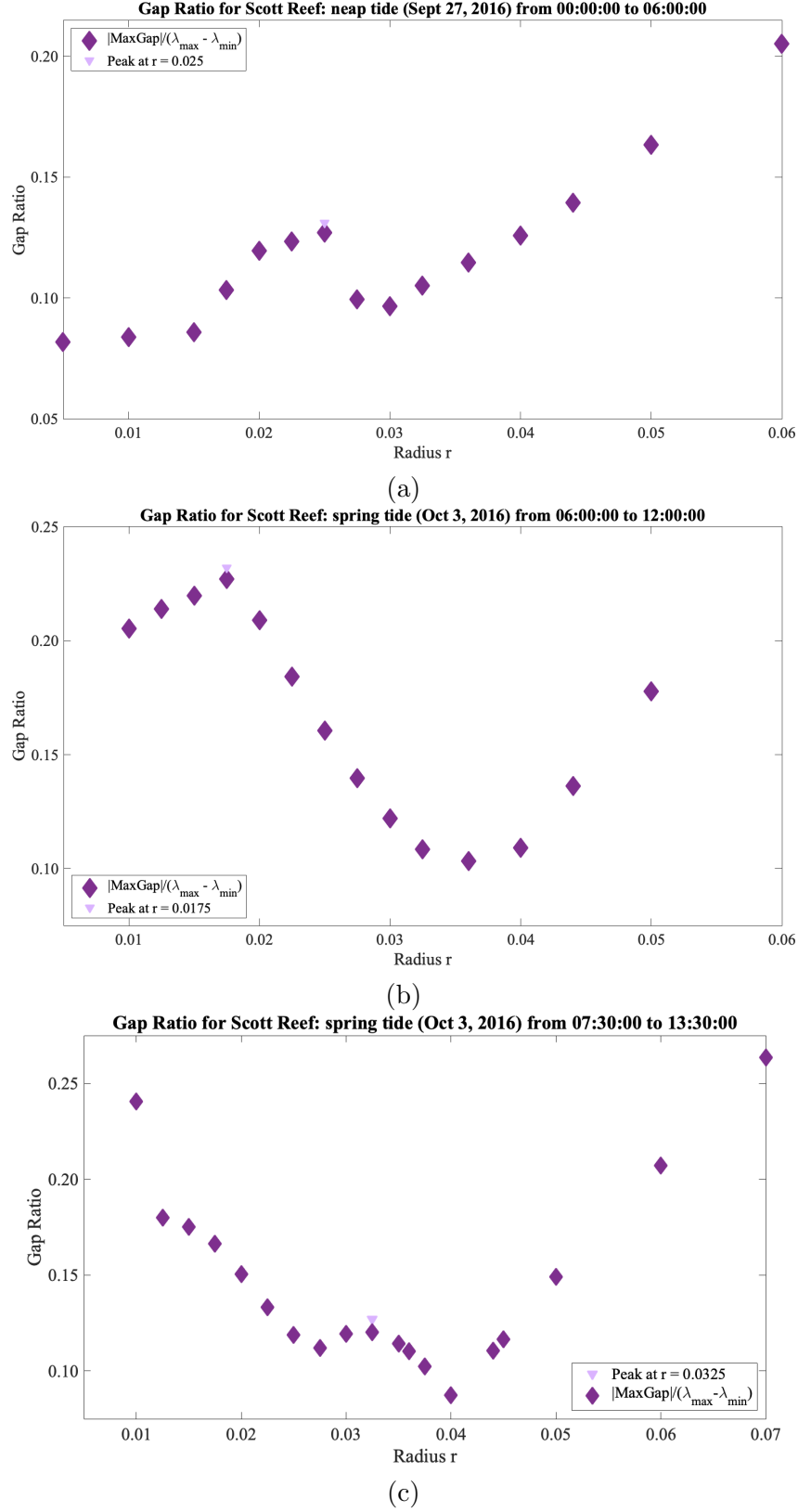


Figure 3-17: Step 1 of the spectral clustering protocol for the Scott Reef channel: gap ratio dependence on radius  $r$ . Sweeps with offset coefficient  $10^7$ . (a) High tide at neap tide. (b) High tide at spring tide. (c) At the start of the drifters release.

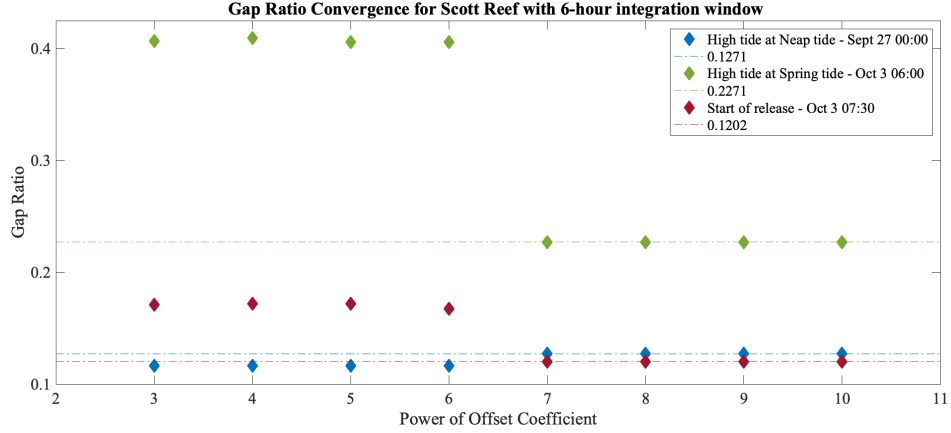


Figure 3-18: Step 2 of the spectral clustering protocol for the Scott Reef channel: gap ratio dependence on the offset coefficient. Sweeps with each of the local maxima from figure 3-17.

### 3.6.4 Ocean physics and the role of surface convergence

The FTLE fields exhibited significant negative values that were of the same order of magnitude as the positive FTLE values. The blue and purple spectral clusters from high tide at spring tide, at 06:00, were located within the areas of highly negative FTLE values. Advection of the spectral clusters also showed that their area shrunk over time. The spectral clustering analysis for the spring tide event was reiterated with a 12-hour integration window to further investigate the shrinkage. The final step of the protocol, the noise coherence calculations, was carried out for the peak in gap ratio with the average distance function, which corresponds to the  $r$  value of 0.01, as opposed to  $r = 0.0175$  for the 6-hour analysis. The detected clusters match the 6-hour results almost exactly and are plotted in figure 3-20.a on page 121.

The previous analysis had shown that after 6 hours, the clusters had considerably shrunk in area; now after 12 hours, they vanished. To investigate the role played by the surface convergence, the shrinkage in cluster area was compared to the vertical fluxes from the divergence of the surface velocity field. The area was calculated in two ways: using the embedded MATLAB *boundary.m* function and using the grid resolution and multiplying it by the number of nodes. The area calculations were validated against polygons computed in QGIS, an open-source geographic information system, as shown in figure 3-20.b on page

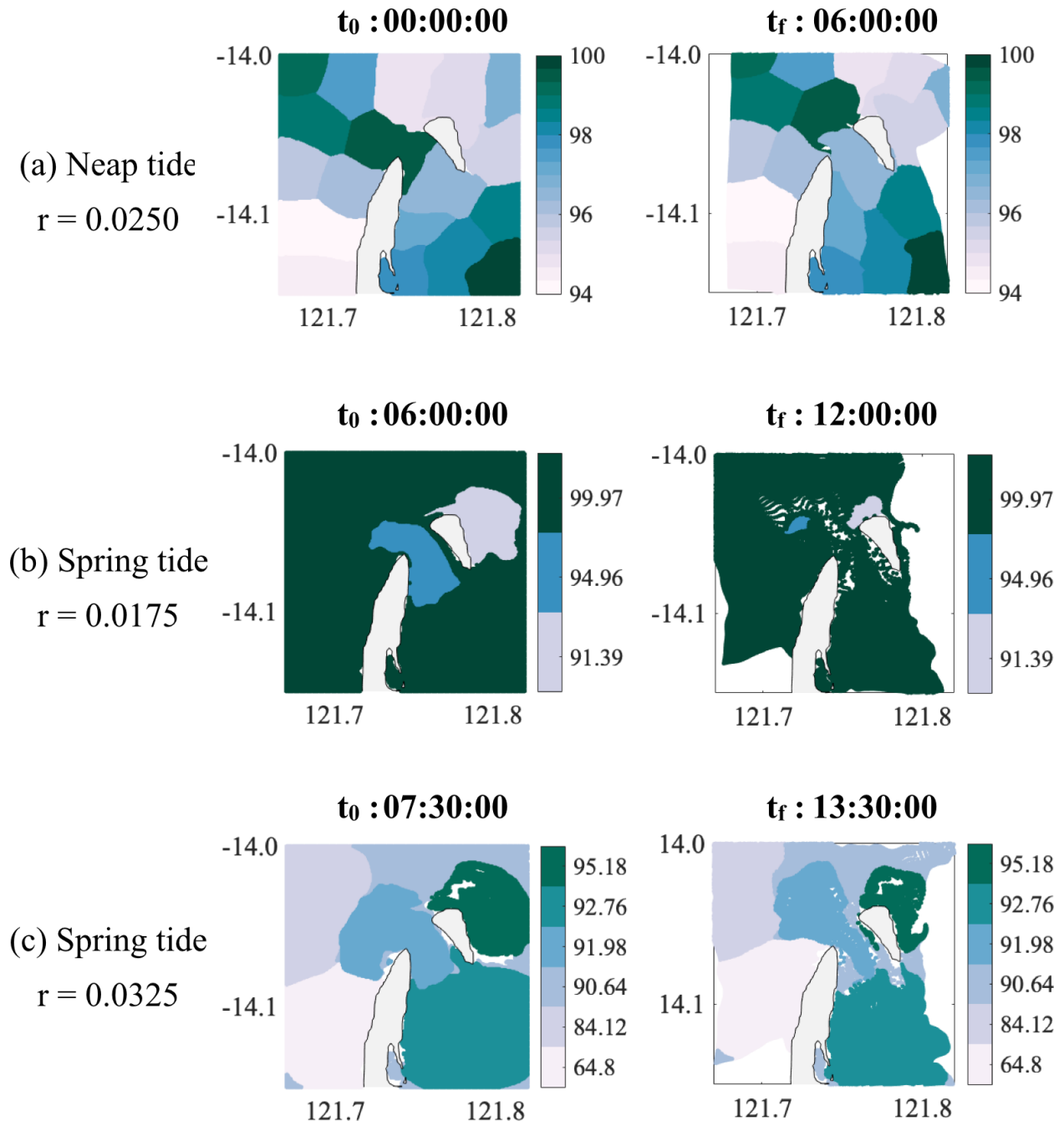


Figure 3-19: Step 3 of the spectral clustering protocol with coherence metrics for the Scott Reef channel. Analysis done on the 2016 dataset (a) at high tide around neap tide on September 27, 2016, (b) at high tide around spring tide on October 3, 2016 (c) at the time at which all drifters were released in water. The  $r$  values correspond to the peaks from figure 3-17.

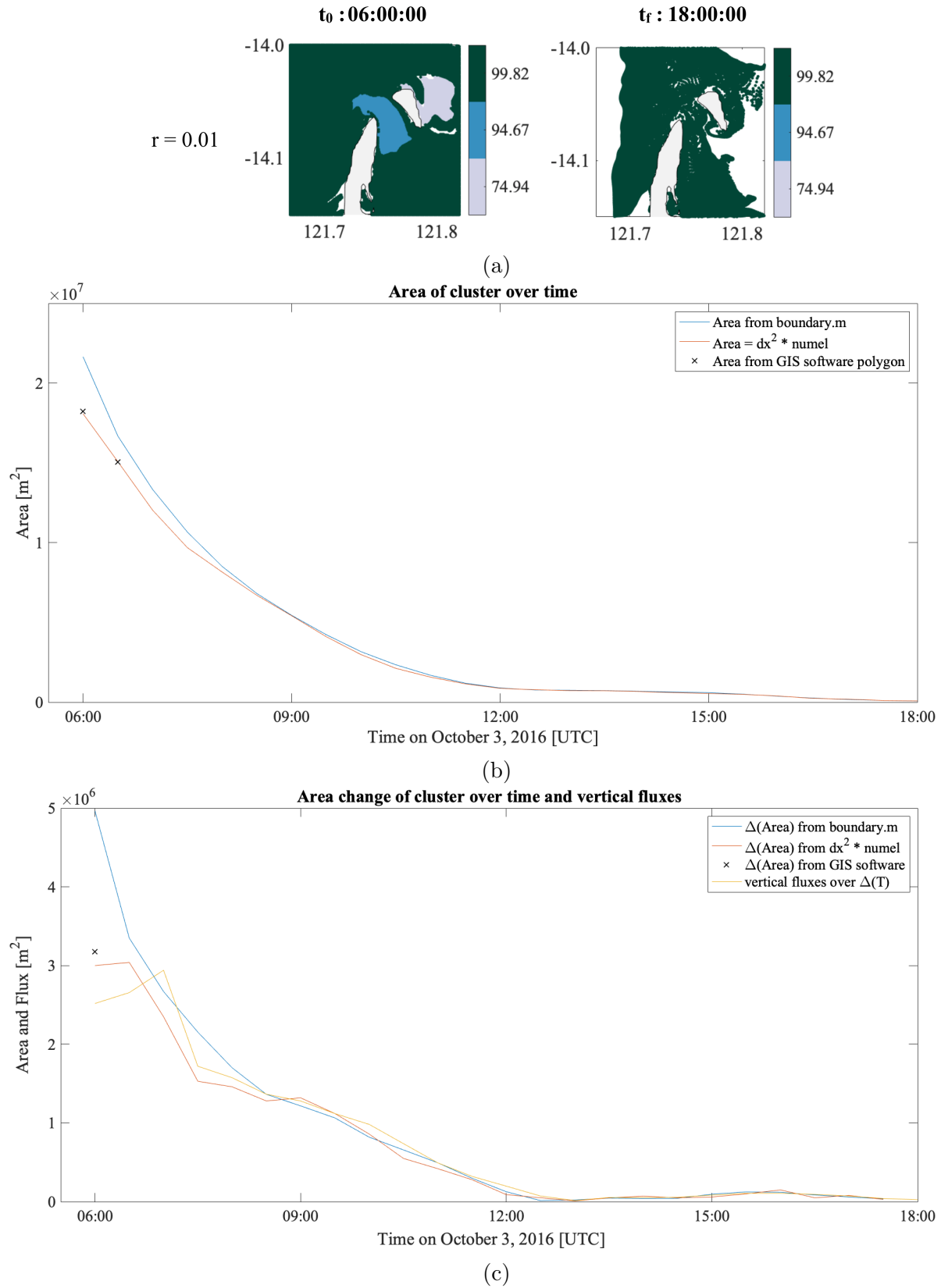


Figure 3-20: (a) Step 3 of the spectral clustering protocol for the Scott Reef high tide event with a 12-hour window. The initial ( $t_0$ ) and final ( $t_f$ ) cluster positions are plotted. (b) Area shrinkage and (c) vertical fluxes for the blue cluster over 12 hours.

121. The vertical fluxes  $\mathbf{F}$  at each time step were measured from the divergence  $\nabla \cdot \mathbf{V}$  of the surface velocities  $\mathbf{V} = (\mathbf{u}, \mathbf{v})$  output by the SUNTANS model, which used the rigid-lid approximation, over the area  $A$  of the cluster:

$$\begin{aligned}\nabla \cdot \mathbf{V} &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = -\frac{\partial w}{\partial z} \\ \mathbf{F} &= \int_T \int_A \mathbf{w} dA \cdot dt dz\end{aligned}\tag{3.1}$$

The comparison is plotted in figure 3-20.c.: the change in area is qualitatively similar to the total vertical flux out of the area. Quantitatively, there are some differences at the first four time steps, but the changes are comparable in magnitude and likely due to numerical resolution. Between 06:00 and 07:30, the area of the cluster is halved, which explains in part why the recalculated clusters at 07:30 were so different from the 06:00 clusters. The vertical fluxes due to surface convergence are also most important before 07:30.

With the anticipation that the cluster's shape was dictated by the surface convergence, the velocity field's divergence component was removed via the least-square estimate method to analyze a non-divergent velocity field during the neap tide event and the spring tide event. The resulting fields, plotted in figure 3-21 on page 123, are virtually divergence-free, with the remaining values being noise. At neap tide, the original velocity fields exhibited divergence one order of magnitude smaller than at spring tide. In both cases, the divergence was maximum inside the channel. At spring tide, the negative divergence was maximum along the channel, somewhat following the shape of the attracting backward FTLE field. This suggests that the high rates of attraction were due, at least in part, to downward convergence within the channel. The two clusters resulting from the parameter-free spectral clustering method were formed around the maximum divergence values, inside the channel and east of Sandy Islet, further suggesting that convergence dictated the spectral clustering results.

The forward FTLE and spectral clustering analyses were reiterated on the divergent-free



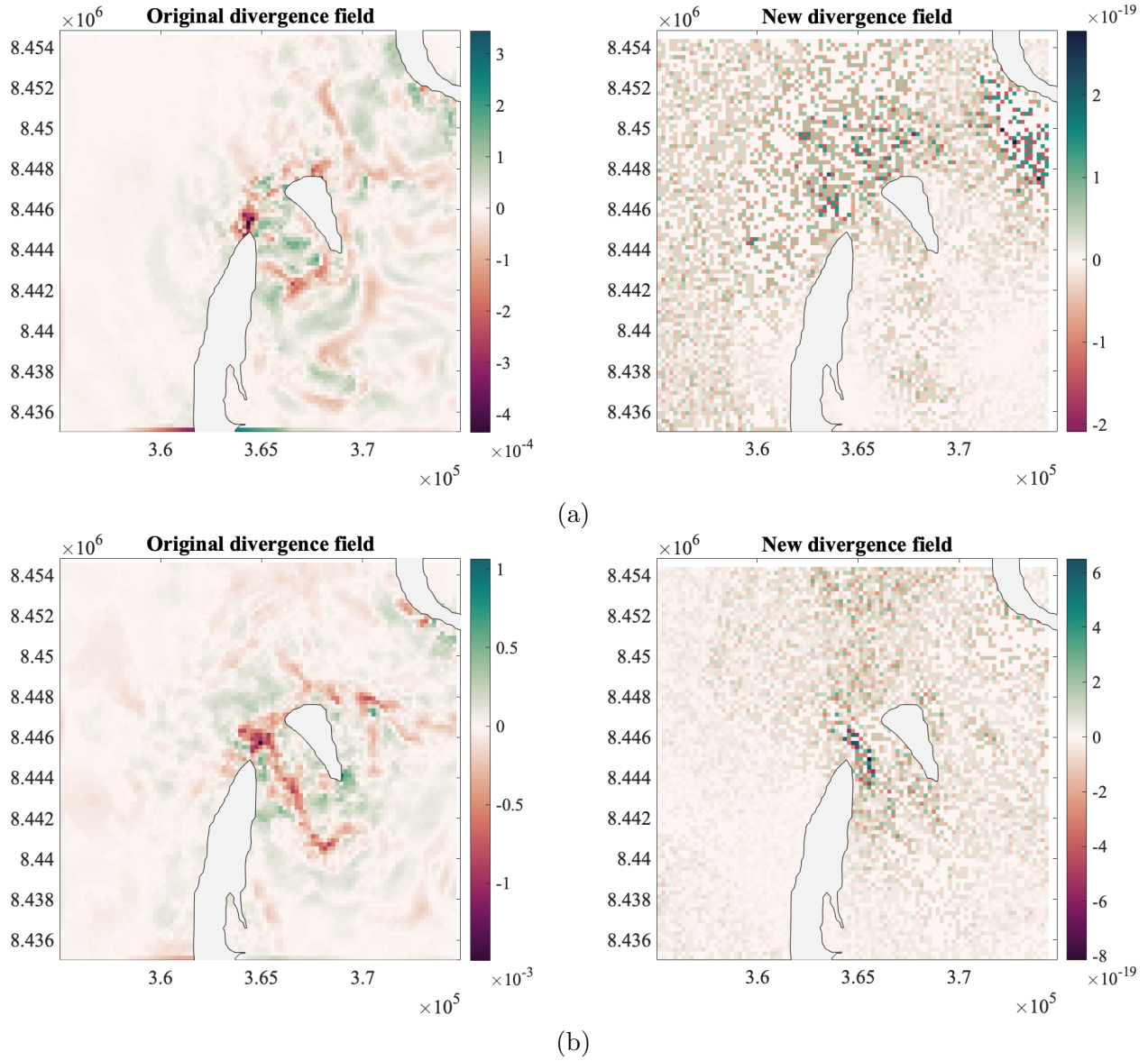


Figure 3-21: Divergence in the Scott Reef channel before and after removing the divergent component via least-squares estimates for (a) the neap tide event and (b) the spring tide event. Coordinates in UTM Zone 51S.

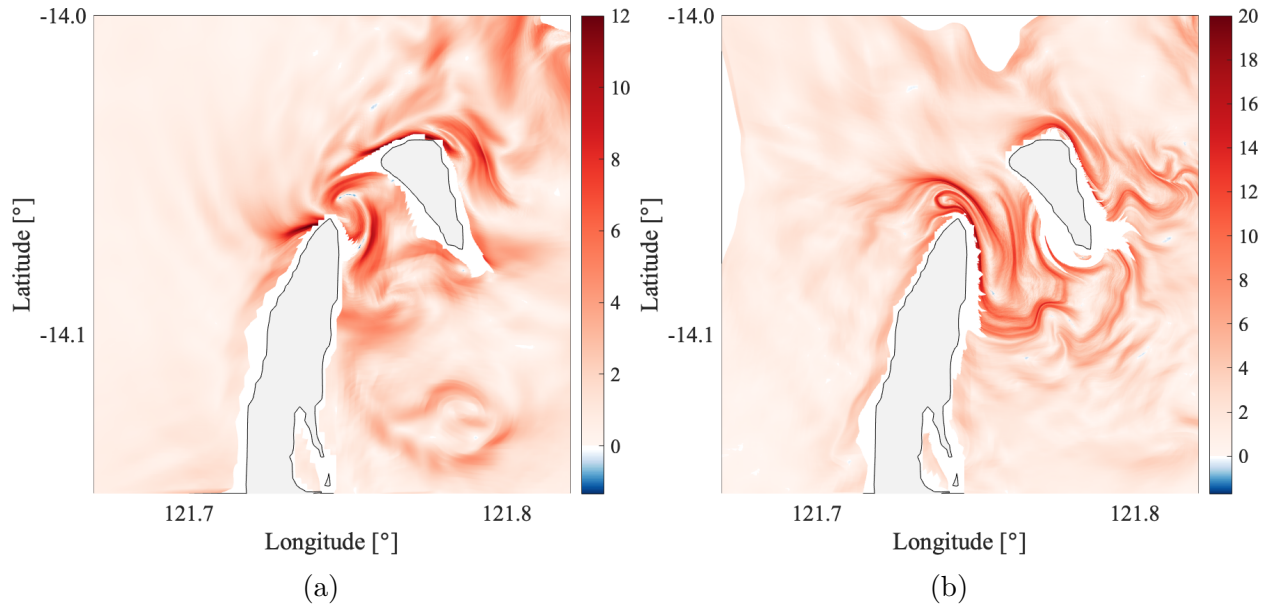
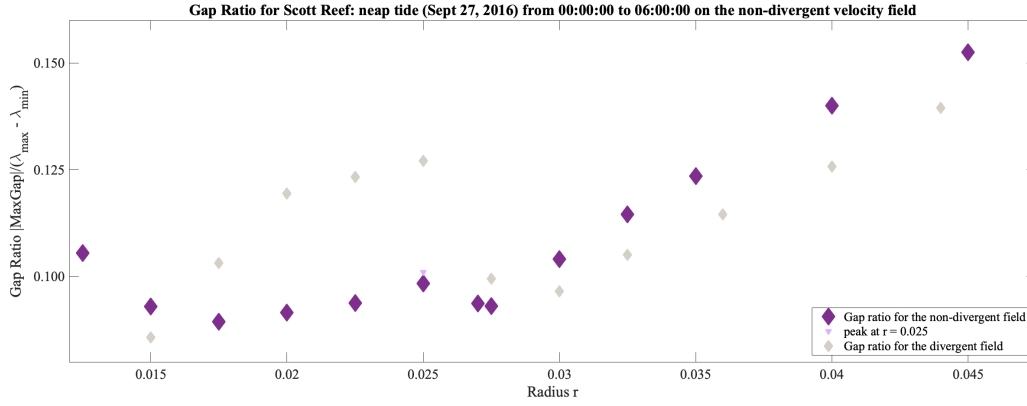
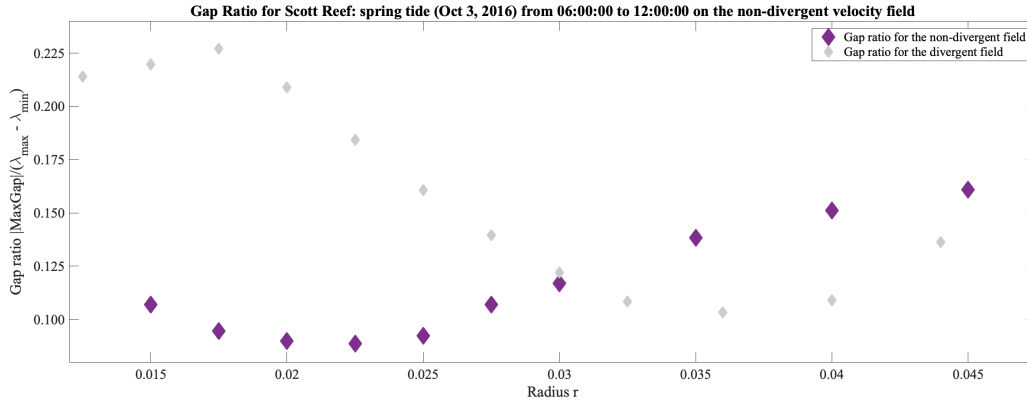


Figure 3-22: Forward FTLE fields in the Scott Reef channel for the non-divergent velocity fields around (a) neap tide on September 27 and (b) spring tide on October 3. The period of integration was 6 hours in all cases.

fields with the same parameters. The FTLE field at neap tide, plotted in figure 3-22.a on page 124, resembles the original FTLE field in figure 3-15a. At spring tide, the pronounced FTLE ridge partially disappears, although high values of FTLE are still found around the tip of Scott Reef South. In both cases, the FTLE values are almost exclusively positive, as they should be for a non-divergent flow. Step 1 of the spectral clustering protocol, the sweep of  $r$  parameters, yielded very similar results for the neap tide event, as shown in figure 3-23.a on page 125. The peak gap ratio occurred at  $r = 0.025$ , same value as for the original, divergent field. This resulted in 18 clusters as well, although their shape and coherent values were a bit different in the non-divergent case, as shown in figure 3-24 on page 125. The similarity in the shape of the gap ratio curve indicates a relative insensitivity to the removal of the divergent component, which is consistent with the relatively small vertical velocity during neap tide as compared to spring tide. For the spring tide event, the shape of the sweep of  $r$  parameters, shown in figure 3-23.b, was very different for the non-divergent field. No peak in gap ratio was detected and the values were much lower: this means that there was no strongly connected component within the domain, unlike in the divergent case in which



(a)



(b)

Figure 3-23: Step 1 of the spectral clustering protocol for the non-divergent velocity fields: high tide during (a) neap tide and (b) spring tide. The original sweeps of parameters corresponding to the divergent flow field is plotted for comparison. At neap tide, the peak occurs at the same value of  $r = 0.0025$ . At spring tide, no peak is detected.

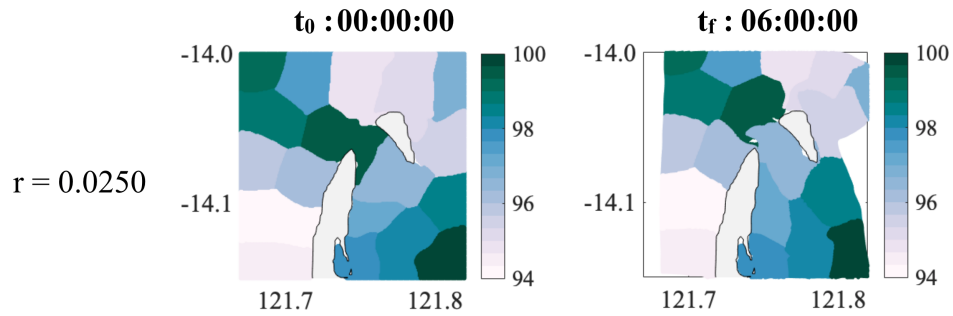


Figure 3-24: Step 3 of the spectral clustering protocol for the non-divergent velocity field at high tide around neap tide: results with coherence metrics for the peak at  $r = 0.025$ .

the two clusters were very distinct from the incoherent background. High tide around spring tide was when the vertical velocities and the divergence values were the highest, one order of magnitude greater than during neap tide.

The removal of the divergent component in the velocity field led to large changes in the FTLE and spectral clustering results. The differences between divergent and non-divergent fields were greater for the spring tide event, which was to be expected as the divergence values were an order of magnitude greater. The key structures identified during the experiment are absent in the non-divergent field. This suggests that the LCS detected during the experiment were to a large degree driven by the three-dimensionality of the flow. The experimental drifters were likely attracted to a convergence zone, but constrained by buoyancy, thus clustering into a small area as the cluster shrunk in size. Over the 07:30 to 13:30 experiment, the ebb tide had turned into flood and the vertical velocities throughout the 6-hour window were much lower than in the 06:00 to 12:00 window, where the tide went from high to low. The influence of the surface convergence also explains why the 06:00 and the 07:30 spectral clustering results are so different.

### 3.7 Discussion

The Lagrangian approach successfully predicted the outcome of the drifter experiment. The FTLE-based predictions using the 2007 model data were able to qualitatively capture the partition of the domain into lagoon waters and offshore waters, predicting that the drifters would split into those staying within the lagoon and those heading out of the channel. Quantitatively, however, the exact location of the predicted repelling ridge did not quite align with the boundary between these two groups of drifters. This is because of multiple reasons: first, the model data dated from 2007, as opposed to the updated 2016 model run; second, the predictions were made for a time of the tidal cycle different from the time at which the experiment occurred; lastly, the drifters may not have been Lagrangian enough. The FTLE

analysis of the 2016 model output led to a much better quantitative agreement between the predicted and observed separatrices, meaning the repelling structure that is the FTLE ridge, dividing the drifters into a southern and a northern group. Drifters released on each side of this repelling ridge exhibited vastly different transport behaviors. The drifter trajectories in the vicinity of the backward, attracting FTLE ridge followed the evolution of this attracting feature. The drifters released within the blue coherent spectral cluster remained a bit more contained, although the inertia of the experimental drifters carried them past the structure. The drifters released within the lagoon waters stayed within the southern cluster. They experienced lower velocities and their momentum did not carry them as far away: the final positions of the numerical particles and the experimental drifters in the southern cluster were much closer. The features detected by the FTLE and spectral clustering methods are not evident in the velocity field, which demonstrates the utility of these methods for such complex flows.

The results from the conventional spectral clustering based on the 2007 model output incorrectly predicted three clusters of drifters: a northern and a southern coherent clusters, as well as an incoherent background. In contrast, the analysis of the 2016 model output using the improved, parameter-free spectral clustering correctly captured the partition of the domain by splitting the real drifters into clusters. The boundary between these clusters aligned well with the forward FTLE ridge. The parameter-free spectral clustering approach worked very well in this example because the gap ratios were quite high for the high tide at spring tide event. The results for the neap tide event are much more inconclusive, especially for the intermediate  $r$  value. This is an example of the limitations of the approach: if the domain is not heterogeneous enough in connectivity, the partitioning struggles to distinguish the strongly connected components from the dissimilar components. This drawback of the method is mentioned in Hadjighasem *et al.* [2017], in which the authors refer to the spectral clustering requirement of having ‘a well-defined spectral gap’ as a ‘weakness’. The absence of strong connectivity within the domain, however, may be a result of interest from

a geophysical point of view, as it means that the flow is simply homogeneous in connectivity. Although the end of the deployment occurred 90 minutes only after high tide, the resulting spectral clusters were quite different from the 06:00 clusters. The most notable difference is the shrinkage of the 06:00 clusters after 6 hours, which was not observed for the two other times.

Further investigations of the physical processes underlying the observed drifter behavior and the shrinkage of the northern cluster suggest that this cluster was largely caused by the strong, downward velocities generated in the channel at spring tide. This is consistent with the strong convergence of surface flow and the associated shrinkage of the northern coherent cluster to near-zero area over the 12 hours. This is further confirmed by the disappearance of the northern coherent cluster in simulations where the divergent component was removed from the surface velocities.

In conclusion, the Scott Reef drifter release was a successful experiment that demonstrated the utility and the applicability of the FTLE and spectral clustering approaches for oceanic surface flows. With the complex bathymetry typical of coral reefs, the FTLE fields identified the high values of separation rates that govern the flow transport. The improved, parameter-free spectral clustering protocol revealed to be helpful and accurate in describing key transient features within the domain.

The LCS at Scott Reef were in large governed by the tidal flow and the resulting convergence at the surface this reversal generates, as seen from the divergence of the flow. Although the vertical velocities were at least one order of magnitude smaller than the horizontal velocities, the resulting divergence led to robust coherent clusters. This was particularly evident for the clusters detected at high tide around spring tide, at 06:00, and was demonstrated by the drifter trajectories, which were constrained by positive buoyancy. This phenomenon, again, was highlighted by the LCS methods, furthermore showing the utility of the Lagrangian approach in understanding the organization of transport.

## 3.8 Acknowledgments

The model used for the study is produced by Matthew Rayson [Rayson *et al.*, 2018]. Alireza Hadjighasem helped with the spectral clustering predictions headed into the cruise. Carlin Bowyer planned and helped the construction of the drifters. Greg Ivey, Ryan Lowe from the University of Western Australia and Thomas Peacock were the principal investigators of the study. James Gilmour from the Australian Institute of Marine Science in Perth was the principal investigator on the cruise, him and the whole crew aboard the research vessel (R/V) Solander made the experiment possible. This work was supported by the National Science Foundation, the Australian Institute of Marine Science, MIT MISTI and I was supported by the Martin Family Fellowship for Sustainability.





# Chapter 4

## Case studies of Lagrangian Coherent Structures around No Man's Land

The NSF-ALPHA project is an international collaborative effort to research the prediction and detection of Lagrangian Coherent Structures (LCS) in oceanic surface flows, with the goal of improving response strategies to search & rescue operations and to pollution events such as emergency oil spills. As part of the project, two series of field experiments were conducted off the coast of Martha's Vineyard, Massachusetts in August of 2017 and of 2018. This chapter focuses on the near-shore experiments that took place around the island of No Man's Land, southwest of Martha's Vineyard.

Section 4.1 presents the scope of the ALPHA project: the motivations behind the efforts and the computational and experimental methods used throughout the experiments. Sections C.1 and C.2 focus on the 2017 and 2018 field work that took place around No Man's Land. The conclusions that can be drawn from the two experiments are synthesized in section 4.4.

## 4.1 Overview

### 4.1.1 The ALPHA project

The Advanced Lagrangian Predictions for Hazards Assessments (ALPHA) project has the goal to develop and test Lagrangian methods for the prediction, mitigation and response to environmental flow hazards, such as oil spills, as well as to help the strategy of search and rescue operations taking place at sea. The project is sponsored by NSF and is an international collaboration led by MIT, WHOI, Virginia Tech, UC Berkeley and ETH Zürich. The project’s approach is multi-faceted and combines theory, computations and experimental observations to implement and utilize Lagrangian methods for oceanic and coastal flows.

Some of the observational barriers [McWilliams, 2016] that are obstacles to the present-day understanding of submesoscale dynamics were covered in chapter 1 of this thesis. To recall, global satellite coverage does not currently offer the spatio-temporal resolution required to follow transient features on the order of hours or days, typical of submesoscale flows. Ship-board instruments offer this resolution, but covering the submesoscale lengths of  $\mathcal{O}(10)$  kilometers within hours is logistically challenging. Rypina *et al.* [2014b] successfully conducted a mass drifter experiment south of Martha’s Vineyard from two vessels using limited manpower, but this type of field study remains infrequent. Within the frame of the NSF-ALPHA project, the feasibility of such submesoscale Lagrangian experiments was investigated through two near-coast field studies with a focus on LCS detection.

To demonstrate and evaluate the predictive capabilities of the Lagrangian approach, several field experiments were launched out of WHOI’s docking facilities and conducted south of Martha’s Vineyard, Massachusetts: two near-shore field experiments taking place over the scales of a few hours and one offshore experiment taking place over a few days. A pilot study was planned in August of 2017 to prepare for the full-scale study of August 2018. In both 2017 and 2018, the two short-scale experiments took place in coastal waters through day-long excursions south of Martha’s Vineyard: one around No Man’s Land, the

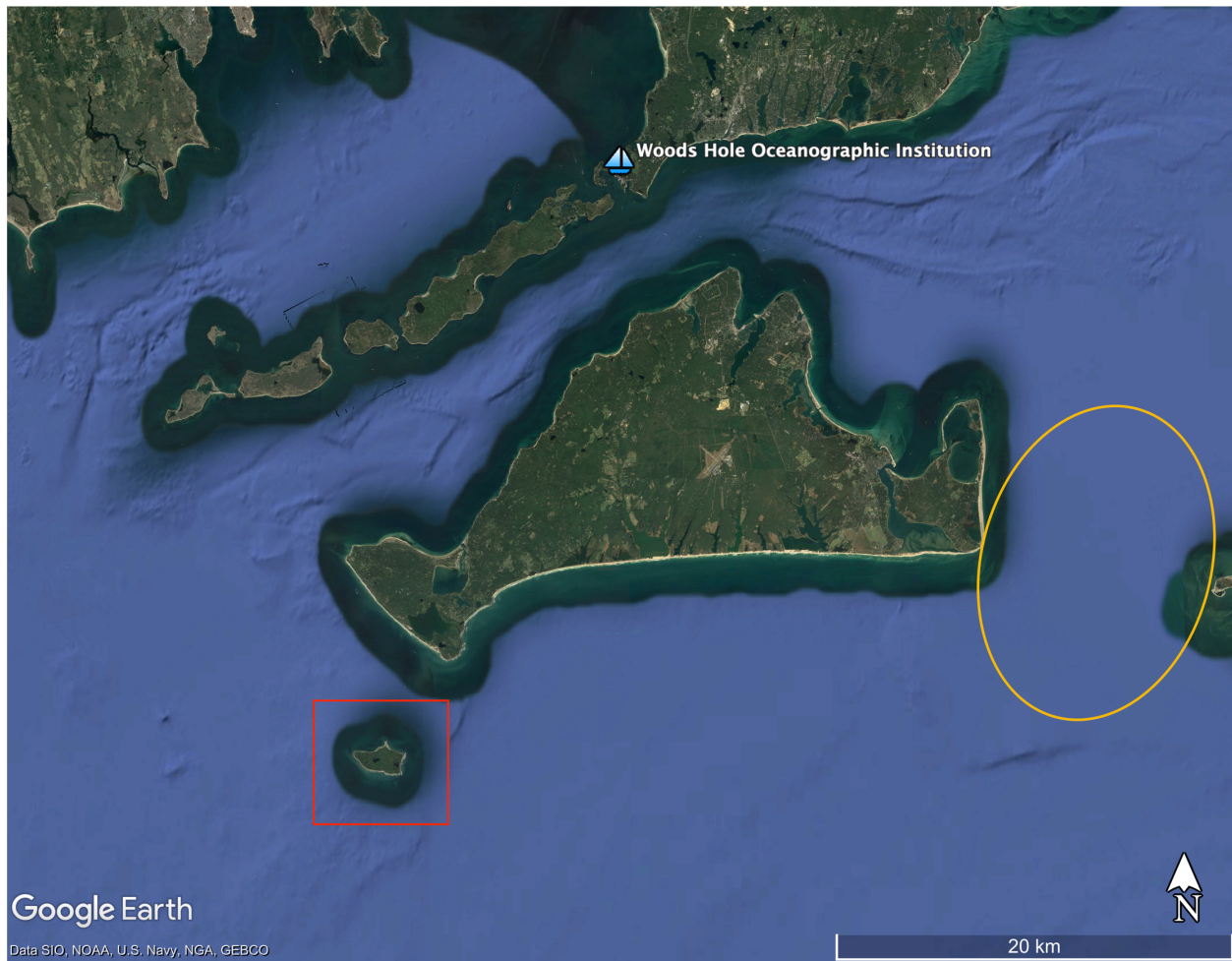


Figure 4-1: Google Earth satellite image of the region of interest around the island of Martha's Vineyard, Massachusetts. The Woods Hole Oceanographic Institution, out of which the field work was based, is indicated by the sailboat. The island of No Man's Land is marked by the red square. The Muskeget channel between Martha's Vineyard and Nantucket is marked by the yellow ellipse.

island marked by the red square in figure 4-1 on page 133, and one in the Muskeget channel between Martha's Vineyard and Nantucket, marked by the yellow ellipse. The focus of the No Man's Land study was the evaluation of Lagrangian Coherent Structures (LCS) as predictive tools to plan and execute field experiments. The Muskeget Channel study focused on the same objective, but for the Objective Eulerian Coherent Structures method by Serra & Haller [2016]. The offshore experiment was conducted over a 4-day research cruise about 30 km south of the Vineyard and focused on vertical shear in velocity and dispersion.

### 4.1.2 Oceanography and bathymetry of the domain

The island of Martha’s Vineyard emerges from the New England shelf [Rypina *et al.*, 2014b], as can be seen in the bathymetry plot of the area is shown in figure 4-2.(a) on page 135. The flow in the region results from wind-driven coastal divergence, tidal forcing and a mean southward drift [Rypina *et al.*, 2014b]. Another feature of interest is a mean westward surface current that reaches velocities of  $0.015 \text{ ms}^{-1}$ , present during the summer months [Rypina *et al.*, 2016]. The bathymetry of the area around Martha’s Vineyard is shown in figure 4-2.b. The Muskeget channel is noticeable by the thin depression in the bathymetry that occurs around  $70.42^\circ\text{W}$ ,  $41.33^\circ\text{N}$ .

Hereafter, this chapter is primarily concerned with the field experiments around No Man’s Land and, particularly, the channel between Martha’s Vineyard and No Man’s Land, which has a width of about 5km. The island is named No Man’s Land because it was used by the U.S. Navy as a practice bombing range until 1996 and is closed to public use for risk of unexploded ordnance. The channel’s depths average 10 meters, with steep bathymetry gradients on each side. Just east of No Man’s Land, the depths increase quickly to 25 meters in less than 2 kilometers.

### 4.1.3 Numerical model

The model used for the NSF-ALPHA project was developed by the Multidisciplinary Simulation, Estimation, and Assimilation Systems (MSEAS) group at MIT using their primitive equation model [Haley & Lermusiaux, 2010; Haley *et al.*, 2015]. The model is based on a two-way nesting configuration: the domain around the continental shelf had a 600-meter resolution and the domain around Martha’s Vineyard had a 200-meter resolution. The two domains are shown in figure 4-2. The models used for the forcing and the data used to initiate the runs are all provided on the MSEAS website: the information concerning the 2017 experiments can be found at [http://mseas.mit.edu/Sea\\_exercises/NSF\\_ALPHA/2017/](http://mseas.mit.edu/Sea_exercises/NSF_ALPHA/2017/) and the information concerning the 2018 experiments can be found at <http://mseas.mit.edu/>

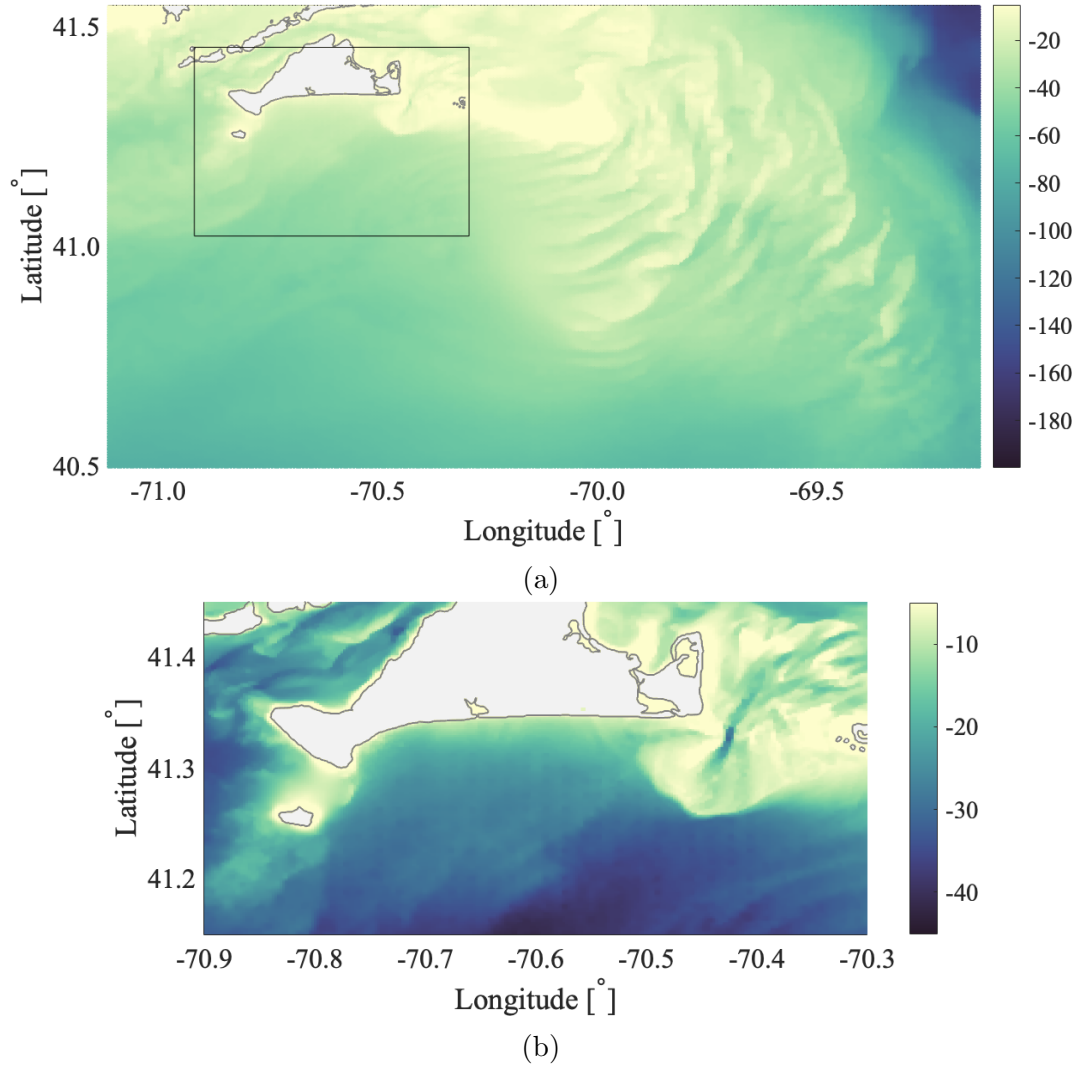


Figure 4-2: Bathymetry of the (a) 600-m and (b) 200-m resolution model domains around Martha's Vineyard, extending south towards the continental shelf break. The domain in (b) corresponds to the rectangle in (a). Depths in meters. Data from MSEAS.

The ocean simulations were forced by atmospheric flux fields forecasts and tidal forcing. To initiate the runs, the models used a variety of ocean and atmospheric observations. Historical data was collected from the National Marine Fisheries Service (NMFS) for conductivity, temperature and depth (CTD). The sea surface temperature data was provided by the Johns Hopkins University’s Applied Physics Lab (JHU APL). For the tidal forcing, the global model of barotropic tide from Oregon State University was adapted to the bathymetry around Martha’s Vineyard. For the atmospheric forcing, the forecasts from the National Centers for Environmental Prediction (NCEP) were used. About a week prior to each field experiment, hydrographic surveys were conducted at sea to collect CTD measurements and initiate the model runs with the current ocean conditions.

Prior to and during the experiment, high-frequency radar data was to be collected from the Martha’s Vineyard Coastal Observatory (MVCO), operated by WHOI, but experimental challenges resulted in a partial coverage only. The HF radar system offers a 400-meter resolution of the 20-by-20-kilometer domain south of the island. The system is described in detail in Kirincich [2016]. Throughout the week of the experiment, a goal was to continuously assimilate measured and predicted environmental data into the MSEAS model. No CTD measurements were available once the experiments started, however, so the data assimilation was limited to a few datapoints of opportunity. All model ensembles were run by the MSEAS team, who also output the model velocity fields. The data was uploaded to the TRACE web-based gateway, described in the previous chapter [Ameli & Shadden, 2019], and can be found at <http://transport.me.berkeley.edu/thredds/catalog/public/MIT-MSEAS/catalog.html>. In this chapter, I used the TRACE web-based gateway for the advection of numerical trajectories, for ease and speed of calculations. The FTLE fields were also computed via TRACE, unless otherwise noted.

#### 4.1.4 Experimental methods

The No Man’s Land experiments consisted of releases of surface drifters targeting the predicted LCS in the vicinity of the island. The drifters used were the same as for Rypina *et al.* [2014b]: the Coastal Dynamics Experiment (CODE)-type, also called Davis type, developed by Davis [1985] and manufactured by metOcean telematics. Pictures of the drifters are provided in figure 4-3 on page 138. The drifters were prepared with the help of Benjamin Hodges from the Autonomous Systems Laboratory at WHOI. For the tracking of the drifter positions over time, a commercially available GPS tracker, the SPOT Tracker model, was used, same model as in chapter 3. The trackers were wrapped in zip-lock bags to protect them from water and attached to their mounts with zip-ties. The mounts had been lathe-turned from plastic pipes to fit the antennas and were secured with electrical tape of a bright color to ease spotting of the drifters during recovery. The additional load on the antennas due to the trackers and their mounts caused a lot of wobbling; thus, to avoid the antennas folding into the water, their bases were reinforced with electrical tape.

The SPOT trackers were set to emit at 5-minute intervals. The trackers - and thus drifters - positions over time were retrievable through the SPOT online portal. To follow the drifters in real-time and to recover them, a MATLAB program was built to automatically retrieve the positions and forward them to the ships via email or text messages. The program retrieved the satellite signals through a Python API from SPOT, downloaded the data onto a computer on shore, then forwarded the newly emitted signals to a cell phone and a computer on the boat.

## 4.2 No Man’s Land 2017 experiment

The first No Man’s Land experiment took place on August 14, 2017. This section describes the predictive analysis used to plan the experiments based on different LCS methods, then presents the experimental trajectories from the drifters and lastly describes the LCS analysis





(a)



(b)



(c)

Figure 4-3: CODE/DAVIS-type drifters. (a) Two drifters stored on deck with the black sails folded. The yellow spheres are hollow floaters for positive buoyancy. The antennas are wrapped with orange reflective tape. (b) A drifter being recovered onboard. The handle was coated with red electrical tape to be more visible and easier to recover. The base of the antenna was reinforced with black electrical tape to minimize its wobbling. (c) A drifter floating. The SPOT tracker was wrapped into a zip-lock bag and zip-tied to its mount. The mount was fixed onto the antenna and reinforced with red electrical tape.



conducted on the updated model run.

### 4.2.1 LCS predictions

Three hydrographic transects were conducted during the week leading to the field experiments to initialize the MSEAS simulations. The numerical velocity fields were output the evening prior to the No Man’s Land experiment, resulting in the predictive LCS analyses being conducted about 12 hours prior to the departure of the ship. The experiments were designed and executed to target several types of LCS: Finite-Time Lyapunov Exponent (FTLE) ridges and Fuzzy C-Means (FCM) clusters. The FTLE method is summarized in chapter 1 and the FCM method is described in chapter 2. The predictions for the day of August 14, 2017 are presented in figure 4-4 on page 140.

The FTLE computations were carried out and output by the MSEAS group using their PDE-based approach. The flow maps were computed for every hour between 10:00 UTC and 22:00 UTC with 6-hour integration windows, a timescale relevant to the tidal cycle. The forward FTLE analysis shown in figure 4-4.(a) for the 14:00-20:00 window of integration predicted FTLE ridges looping around No Man’s land and inside the channel between the eastern tip of No Man’s Land and the southern tip of Martha’s Vineyard. These ridges correspond to the curves along which the rates of flow particle separation are local maxima and are referred to as repelling structures. The predicted ridge formed in the channel separates the waters east of the channel and south of Martha’s Vineyard from the waters west of the ridge. The backward, repelling ridge became progressively more pronounced between the time window of 17:00 to 11:00 and the time window of 20:00 to 14:00, after which the structure disappeared. For this reason, the drifter deployment was planned for 14:00. The backward FTLE fields in figure 4-4.(b) predicted a strong attracting ridge west of Martha’s Vineyard, meaning that the material in the vicinity of this ridge would get attracted northward. Strong rates of particle attraction were also predicted west of No Man’s Land. Due to the complex bathymetry and topography, the flow dynamics around the coast

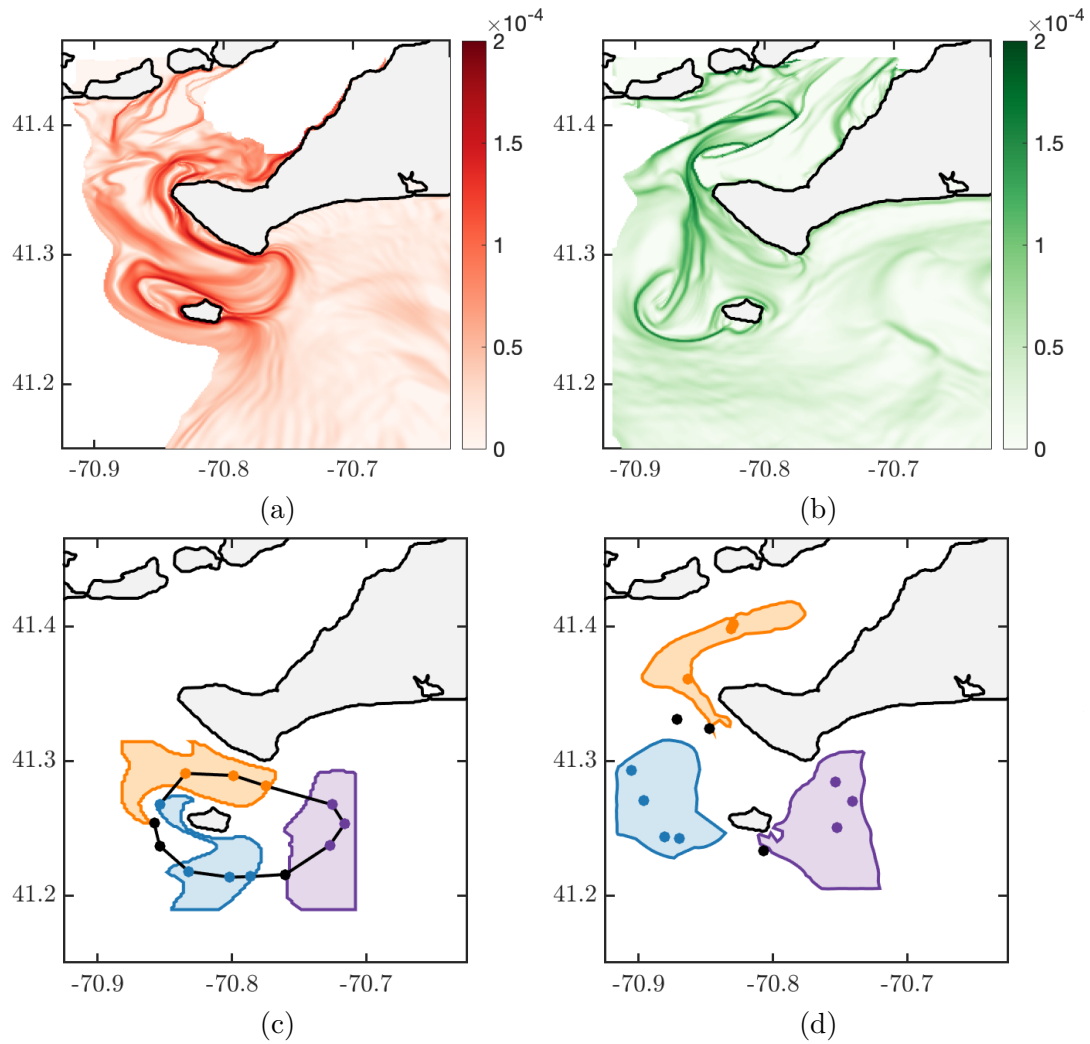


Figure 4-4: LCS predictions used for the 2017 No Man's Land experiment. (a) Forward-time 6-hour FTLE field computed by MSEAS from 14:00 to 20:00. (b) Backward-time 6-hour FTLE field computed by MSEAS from 20:00 to 14:00. (c) FCM clusters computed over 6 hours from 14:00 to 20:00 by Michael Allshouse. The dots correspond to the suggested release locations. The colors correspond to the FCM cluster in which the drifters would fall. (d) Location of the FCM clusters and the drifters at the final time of 20:00.

are particularly arduous to capture. High values of FTLE were found along the coastlines and at the boundaries of the domain in both forward and backward times: these high rates of separation or attraction of numerical particles may be due to the limits of the computation domains, to advection of material onto shore or other coastal effects.

The FCM analysis was conducted by Michael Allhouse at Northeastern University for the 14:00-20:00 time window and is shown in figure 4-4.(c). Three FCM clusters were predicted around No Man's Land: an orange cluster inside the channel, a blue cluster south of No Man's Land and a purple cluster east of No Man's Land. The final positions of these clusters are shown in figure 4-4.(d), where the orange cluster was advected North, the blue cluster was advected west of No Man's Land and the purple cluster stayed east of No Man's Land. The predictive FCM results are shown for the cores of the clusters only, as the clusters were filtered by probability of membership for the prediction. This means that only particles with a high enough likelihood of belonging to their respective clusters were kept for this figure.

Based on the FTLE fields and the FCM clusters, the proposed release strategy was to deploy drifters in an elliptical configuration around No Man's Land, which is plotted in figure 4-4.(c) by the dark line. The assumption was that the colored dots would be released within their respective clusters: orange, purple or blue. Some drifters would be deployed outside the cores of these clusters, marked in black dots. The projected final positions of these drifters are shown in figure 4-4.(d). For the deployment, plan was to follow the ellipse in the clockwise direction starting northwest of No Man's Land, so that the first drifters would be released west of the forward FTLE ridge. By the time all drifters are released, a new ridge would be formed west of No Man's Land and the latest drifters would be released east of this ridge. The expectations were that the drifters would end up into one group south of No Man's Land and two groups on each side of the ridge, one east and one west, with the western drifters ending up north.

### 4.2.2 Experimental results

The experimental trajectories of the CODE drifters are presented in figure 4-5 on page 143. The first drifter, denoted by number 1 in the figure, was released at 14:20 UTC and the last drifter, denoted by number 14, was released at 15:51 UTC, meaning that the deployment of drifters lasted over 1.5 hours out of the 6-hour window of integration used for the predictive analysis. On this day, low tide occurred at 15:02, so all drifters were released at a similar point of the tidal cycle. The rate at which the drifters were released in the water was limited by the vessel's capabilities and by the deployment of the individual drifter: the vessel needs to slow down around the targeted station in order to release the CODE drifter in water. The trajectories are plotted until 21:51 UTC, 6 hours after the last drifter was in water. The trajectories are color-coded in time to show the evolution of the drifters, as well as the lag between the times at which the first and last drifters were released.

The drifter trajectories indeed seemed to split into three qualitatively different behaviors: two drifters headed north, seven drifters experienced a reversal in heading and turned eastward past 18:30 and five drifters were advected west throughout the release. Specifically, drifters 2 and 3 were released west of the predicted repelling ridge and were advected northward past the western tip of Martha's Vineyard. Drifters 4 through 9 were released east of No Man's Land; they first were headed west and turned around at about 18:30 UTC. Drifters 4 through 6 in particular were released within the channel then followed the coastline while heading east. Drifters 10 through 14 were released south of No Man's Land and were advected westward throughout the release without experiencing any reversal in direction. Drifter 1 started being advected northward then looped around with the tide. Two triplets were released: drifters 4 through 6 and drifters 12 through 14.

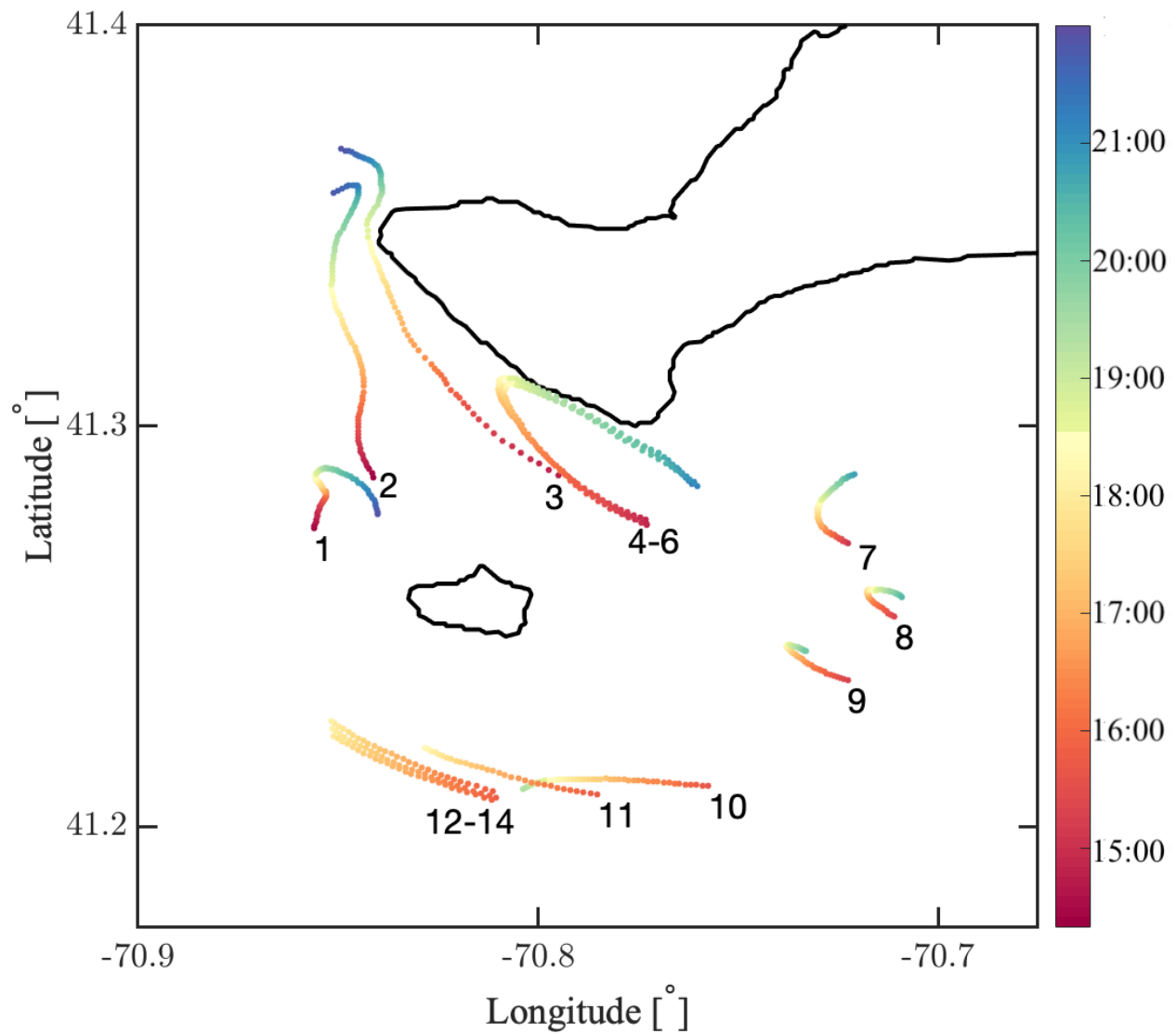
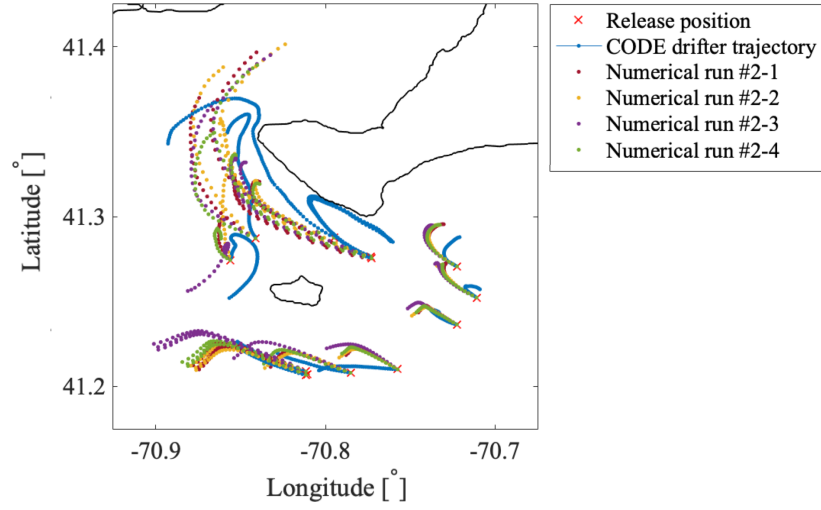


Figure 4-5: CODE drifter trajectories for the August 14, 2017 No Man's Land experiment. The drifter positions are color-coded by time.

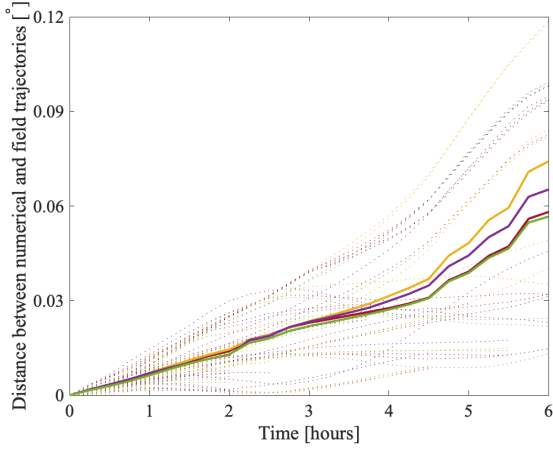
### 4.2.3 Comparison of experimental and numerical trajectories

Several numerical ensembles were output by the MSEAS model to represent different initializations, as the forecasts contained uncertainties. For the day of August 14, 2017, the day of the No Man’s Land experiment, four ensembles were output for the 200-m grid. The numerical trajectories were computed from the MSEAS velocity fields with the TRACE portal. Numerical tracers were seeded in the four different velocity fields, on an individual basis to start the numerical advection with the corresponding drifters’ initial positions and release times. The results are shown in figure 4-6.(a) on page 145, in which the experimental trajectories from the CODE drifters are plotted in thick blue lines and the numerical trajectories are plotted in colored dots: red, yellow, purple and green for ensembles 1 through 4, respectively. The key differences between model trajectories and CODE trajectories were found for drifters 4 through 6: this triplet looped back eastward along the coast, whereas the numerical trajectories continued being advected northward. This disagreement between model and experimental trajectories was not surprising: this triplet was released in the channel between No Man’s Land and Martha’s Vineyard, in a sensitive region of the domain where the coastal dynamics were hard to capture. Their exact location relative to the repelling LCS was uncertain and the fate the trajectories varied widely depending on which side of the repelling ridge they were released.

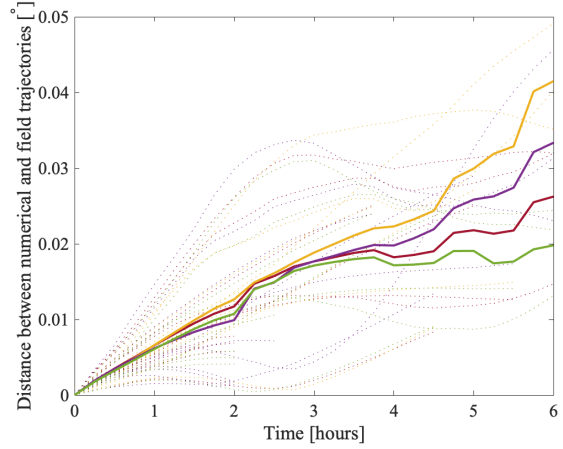
To pick the ensemble among the four that most faithfully represented the behavior of the CODE drifter trajectories, the pairwise distances between the experimental and numerical trajectories were calculated, for each ensemble and each drifter. The results are shown in figure 4-6.(b). On average, the smallest deviations between experimental and numerical trajectories were found for ensemble 4, marked with the green thick line, although all runs gave similar results. When omitting drifters 4 to 6, the triplet of drifters that turned around while their numerical counterparts continued being advected in the channel, the distances were considerably smaller on average and ensemble 4 performed noticeably better than the other ensembles, as shown in figure 4-6.(c). This ensemble was thus the one chosen to proceed



(a)



(b)



(c)

Figure 4-6: (a) CODE drifter trajectories (thick blue lines) and numerical trajectories individually seeded for each ensemble of the model (colored dotted lines). The release positions are shown in red crosses. (b) Pairwise distances between the experimental and numerical trajectories, for each ensemble and each drifter. (c) Same as (b) minus drifters 4-6.

with the more in-depth LCS analysis, described in the next section.

The velocity fields were output for the 14:20 to 20:20 UTC time-window to show flow transport features when the first drifter was released and for the 15:51 to 21:51 UTC time window to show flow transport features when the last drifter was released. For the clustering methods, the 15:51 outputs are also shown at 21:51 UTC, six hours later, as all methods were run with a 6-hour window of integration. For the FTLE and encounter volume methods, the velocity fields had to be output for the time window from 21:51 UTC on August 14 to 03:51 UTC on August 15, thus overlapping two different days. The model run’s time window finishing at 00:00 on August 15, the output of a different model run was used for these specific cases. The fourth ensemble was chosen as well, but the initialization for the run was slightly different.

#### **4.2.4 LCS analysis**

The numerical model data was processed with four LCS methods: FTLE, encounter volume, FCM and the parameter-free spectral clustering from chapter 2. The results are plotted in figures 4-7–4-8 on pages 147–151. Figure 4-7, part 1, shows the forward and backward FTLE fields in panels (a) and (a’), respectively. The encounter volume results are shown in panel (b). The FCM clusters are shown in panel (c) and the spectral clusters in panels (d-d’). A zoom in on the clustering results at 21:51 is shown in figure 4-8 on page 151. The CODE drifter positions are superimposed to the results. Appendix C provides additional details about the results of the FTLE, the spectral clustering and the encounter volume methods.

##### **4.2.4.1 Finite-Time Lyapunov Exponent (FTLE)**

The forward (figure 4-7.a) and backward (figure 4-7.a’) FTLE fields were computed by the TRACE portal to look at repelling and attracting structures. The FTLE method is included here because was used for the predictive analysis and it is the most established LCS method. The drifter markers are changed according to the side of the FTLE ridge the drifters were re-



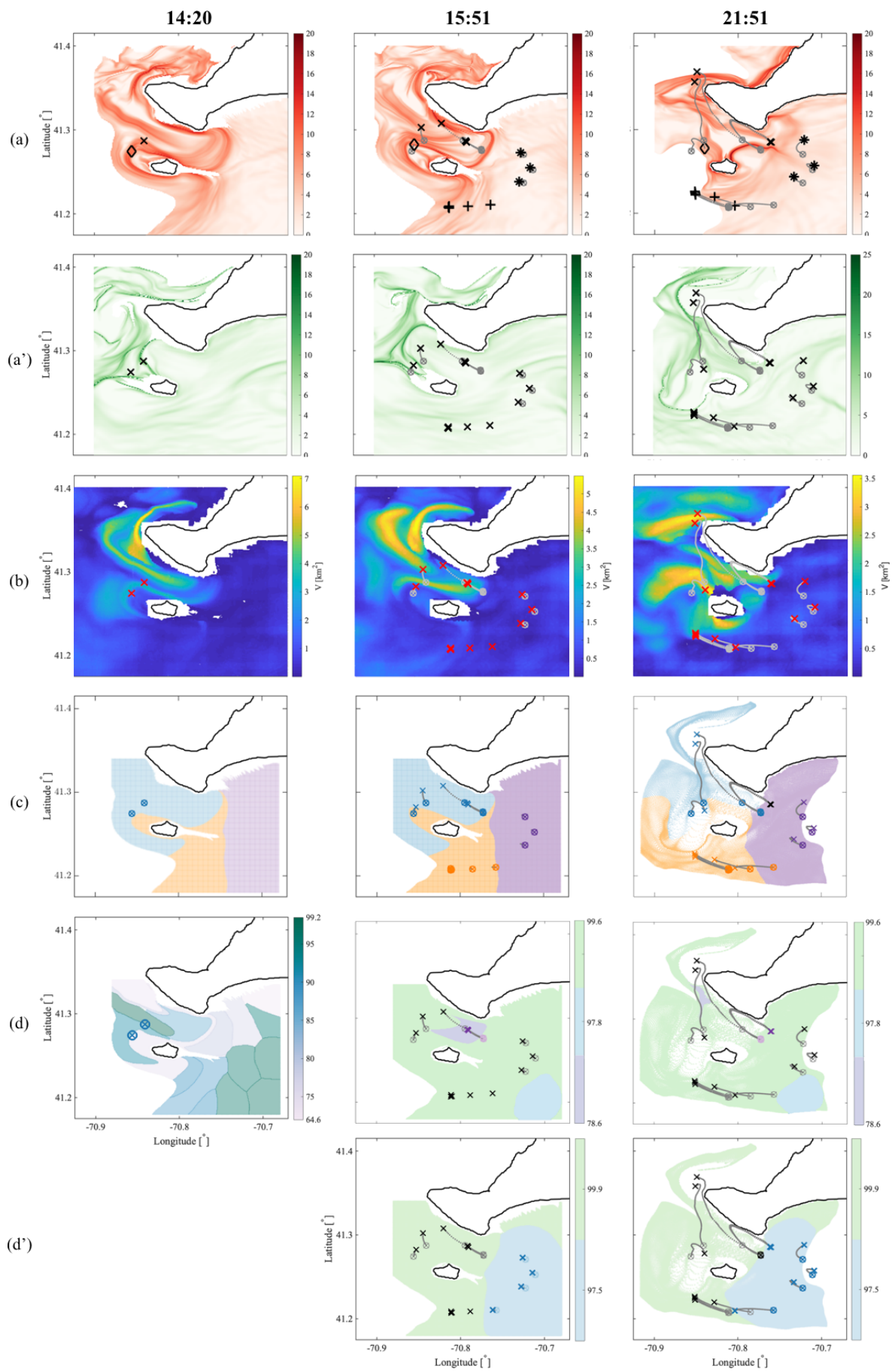


Figure 4-7: LCS and drifter positions for the 2017 study. (a) Forward and (a') backward FTLE (see section 4.2.4.1). (b) Encounter volume (see 4.2.4.2). (c) FCM clusters (see 4.2.4.3). (d) and (d') Spectral clusters with coherence metric (see 4.2.4.4 and figure 4-8).

leased on. The FTLE analyses of the pre- and post-experiment numerical model data yielded similar results and the results did not change much between 14:20 and 15:51. Appendix C presents the FTLE fields in forward time at different times of the tidal cycle for August 14, 2017, further showing that the fields were qualitatively similar between 14:00 and 16:00.

In forward-time, the field shows the rates of separation being neighboring particles, with high values of separation in dark red. A strong repelling ridge occurred east of the channel at 14:20. At 15:51, a ridge was formed slightly more to the west. At both times, very high values of forward-time FTLE occurred around the western coast of Martha's Vineyard and around No Man's Land. It can be attributed to the high rates of flow separation around the coastlines, likely due to the bathymetry. The two ridges were connected from the western tip of No Man's Land to the southern tip of Martha's Vineyard, indicating a separation between the waters east of the ridges, south of the Vineyard, and the water west of the ridges. East of the ridges and south of the Vineyard, the FTLE fields exhibited very low values. At 21:51, high values of forward-time FTLE were found off the eastern tip of No Man's Land, but not connecting to the other coast of the channel.

Drifter 1, marked with a diamond, was released inside the FTLE loop around No Man's Land; it started being advected slightly north then looped around to the south. Drifters 2 through 6, marked with crosses, were released west of the 14:20 and 15:51 ridges. Drifters 2 through 3 were advected northward, whereas drifters 4 through 6 were released inside the channel, slightly west of the ridge, then looped around to the east. Drifters 7 through 9, marked with asterisks, were released east of the ridge; they started heading west then looped around towards to east. By 21:51, the previous repelling structure had disappeared, with high values east of No Man's Land, but no ridge bridging the channel to Martha's Vineyard. The difference in transport behavior between the drifters released inside the channel and the drifters released east of the channel was likely due to the presence of the repelling FTLE ridge. Drifters 10 through 14 were released south of No Man's Land; their heading was consistently westward, exhibiting a transport behavior different from the previous groups of

drifters.

The backward-time FTLE indicates the rates of attraction between particles, with high values of attraction in dark green. The field showed high values of attraction northwest of Martha's Vineyard at 14:20 and at 15:51. For the integration window from 21:51 to 15:51, sharply defined ridges appeared west of No Man's Land and of Martha's Vineyard. Two drifters were advected northward, their trajectories following the ridges, as shown in figure 4-7.a'. Most drifters were released in areas with low backward FTLE values.

#### **4.2.4.2 Encounter volume**

The encounter volume was computed to look at mixing potential around the domain of interest. The encounter volume, typically, requires a larger computational domain to allow particle trajectories to be advected on length scales large enough to give meaningful results. It was thus calculated from the fourth ensemble's output as well, but for the 600-m grid. The results are cropped to the domain of interest in figure 4-7.b, but the full results are shown in appendix C.1.3. The encounter volume method was run with radii from 0.25 to 10 km. Convergence in the number of encountered trajectories was reached below radius 0.5 km. The 0.5 km results are thus plotted in the figure. At 14:20 and 15:51 the highest rates of mixing potential were found west of Martha's Vineyard, roughly along the backward, attracting FTLE ridges. At 21:51, the highest rates of mixing potential are found northwest of No Man's Land and may indicate convergence.

#### **4.2.4.3 Fuzzy C-Means (FCM)**

The FCM clusters are plotted in figure 4-7.c, with each cluster being indicated by a different color. The drifters are color-coded according to the cluster they were released in and/or the cluster they ended up on; the initial positions are shown with circles and the final positions are shown with crosses. The trajectories are plotted with grey dots. If the drifters were not released in any cluster, their positions are shown in black. The FCM output displays

the geographical partition of the domain. It is one of the methods chosen here because it was used for the planning of the experiments. The method was run with  $K = 3$  to match the parameter setting chosen for the predictive analysis. All drifters were released within a cluster: blue, orange or purple. Drifters 1 through 6 were released within the blue cluster. Drifter 1 was released at the boundary between the blue and the orange cluster and stayed within the blue cluster at all times. Drifters 4 through 6 *left* the blue cluster to turn around and end within the purple cluster. Drifters 7 through 9 stayed within the purple cluster, although they turned around eastward before the numerical tracers did. Drifters 10 through 14 stayed within the orange cluster.

#### 4.2.4.4 Parameter-free spectral clustering

The parameter-free spectral clustering approach from chapter 2 was used to look at connectivity. It is shown in 4-7.(d-d'), where each row corresponds to different peaks in the gap ratios. The sweeps of  $r$  radii are shown in appendix C.1.2. The spectral clustering results at 14:20 indicated a domain that was fairly homogeneous in connectivity, as it was partitioned into multiple adjacent clusters. The maximum gap ratio was found for a radius of  $r = 0.020$ . The coherence metric indicated very coherent clusters northwest of No Man's Land and at the southeastern part of the domain. At 15:51, two peaks in gap ratio occurred, at  $r = [0.0125, 0.0425]$ . The first peak yielded two coherent clusters: a purple one with a 78.6% coherence percentage and a blue one with a 97.8% coherence metric. The second peak yielded one large cluster with a coherence value of 97.5%. For the first 15:51 peak, drifters 4 to 6 were released within the purple cluster before looping around. A zoom-in on the final positions is shown in figure 4-8. For the second 15:51 peak, drifters 7 to 10 were released within the blue cluster and stayed throughout the release, with drifter 10 ending at the border between the coherent cluster and the incoherent background. Drifters 4 through 6 were released within the incoherent domain, here, then ended in the blue cluster.

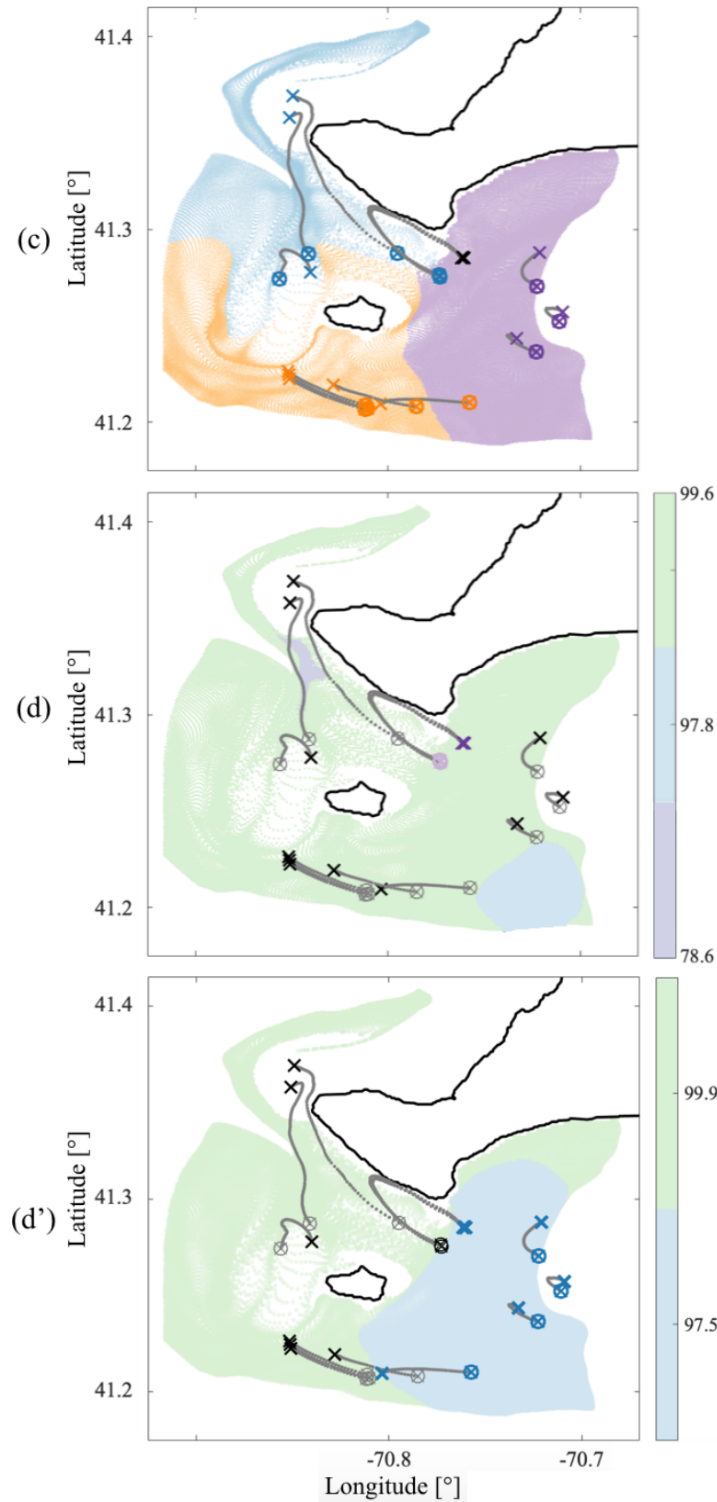


Figure 4-8: LCS and drifter positions for the 2017 study: zoom in on the results at 21:51, 6 hours after all drifters were in water and at the final time of integration of the LCS analyses. The initial and final positions of the drifters are indicated by a small circle filled with a small cross and a large cross, respectively, and color-coded according to the cluster in which they were released. If they ended in a different cluster and/or if they were in the incoherent background, the markers are black. (c) FCM clusters, each indicated by a different color. (d-d') Spectral clusters with coherence metrics at each peak in gap ratio.

### 4.2.5 Discussion

The four LCS methods provided different types of information about flow transport. The forward FTLE ridge, the FCM and spectral clusters indicated that the channel between No Man’s Land and Martha’s Vineyard separated the domain into two parts: waters east of No Man’s Land that stayed south of the Vineyard and waters west of No Man’s Land. This eastern subdomain corresponded to the purple FCM clusters and the blue spectral cluster. Connectivity was found to be the highest in the southeastern part of the domain. This large connected subdomain also exhibited very low FTLE values, in both forward and backward time, meaning low rates of particle separation, and low encounter volume, indicating a low mixing potential. The combination of these outputs indicates that the waters east of No Man’s Land and south of Martha’s Vineyard do not flow out of the channel and do not mix well with the waters outside of the channel. This region may be constrained by bathymetry: as seen in figure 4-2, the depths within the channel are slightly shallower than the depths south of the Vineyard, with a steep gradient.

The backward FTLE ridge, the purple spectral cluster at 15:51 and the blue FCM cluster indicated that material northwest of No Man’s Land was advected northward, west of Martha’s Vineyard. This was clearly confirmed by the CODE drifters 2 and 3. The disappearance of the purple spectral cluster with the second peak in gap ratios is likely due to the  $r$  radius, which went from 0.0125 for the first peak to 0.0425 for the second peak, and was too large for this purple cluster to appear.

The orange FCM cluster seemed to correspond to the incoherent background from spectral clustering, although there is no way of knowing from the FCM cluster partition. Drifters trajectories 10 through 14 stayed within this cluster throughout the release. What makes this group of drifters interesting is that they did not experience a reversal of directions, unlike all other drifters which were released north of them. This confirms the existence of a different transport behavior.

The CODE drifter trajectories were mostly in agreement with the LCS analysis. The

obvious exceptions are drifters 4 to 6, which left their assigned clusters both with FCM and with spectral clustering. These numerical trajectories were also the only ones exhibiting transport behaviors vastly different from the CODE trajectories. For all other trajectories, the drifters stayed within their assigned clusters. Note, however, that the trajectories of the CODE drifters were fairly short in comparison to the size of the clusters: while they do not contradict the FCM partition, the trajectories did not cover lengths big enough to leave the clusters and invalidate the LCS analysis.

### 4.3 No Man’s Land 2018 experiment

The main, full-scale ALPHA experiments took place in Summer 2018. The study area was still No Man’s Land, building on the preliminary results from the 2017 pilot survey. The 2018 No Man’s Land experiment took place on August 7. One goal of the 2018 experiment was to deploy drifters over a larger domain than in 2017. A similar elliptical route around No Man’s Land was used: in 2018, it extended over 20 km, as opposed to 16 km in 2017. The 2017 drifter deployment having lasted over 1.5 hours out of the 6-hour time window of interest, two vessels were used for the 2018 experiment with the aim to speed up the deployment. The ellipse was split in a northern and a southern part to minimize ship time. The plan was to start deploying early so that by the time the LCS of interest would be formed, all drifters would be in water. The number of drifters increased from 14 drifters at 10 locations in 2017 to 18 drifters in 2018. This section describes the coherent structures that were calculated and used to plan and execute the field study, then presents the experimental results and the analysis conducted *a posteriori*.

### 4.3.1 LCS predictions

Similarly to 2017, a few hydrographic transects were completed the days prior to the field experiments to initialize the MSEAS simulations. The numerical velocity fields were output the evening prior to the No Man’s Land experiment, resulting in the predictive LCS analyses being conducted about 12 hours prior to the departure of the ship. Two LCS methods were used in the planning and execution of the experiment: the FTLE fields, output by MSEAS, and the Objective Eulerian Coherent Structures (OECS), according to the method by Serra & Haller [2016], both shown in figure 4-9 on page 155. The OECS method is not part of this thesis, as it is not a Lagrangian method, and is only mentioned here as it was part of the planning of the experiment.

The FTLE fields were computed from 16:00 to 22:00. The forward-time field resembled the picture from 2017: a pronounced repelling ridge was formed east of the channel, connecting the eastern tip of No Man’s Land to the Southern tip of Martha’s Vineyard. The backward-time FTLE showed an attracting ridge formed just west of the repelling ridge: this ridge connected the eastern tip of No Man’s Land to Martha’s Vineyard as well, all along the coast up until the western tip of the island. This attracting ridge along the coast is likely due to material being advected onto shore and/or due to the complex flows occurring along the coast.

The OECS fields were calculated by Mattia Serra at 14:00. Briefly, OECS can be thought of as the instantaneous limits of the FLTE and the purpose of this analysis was to seek high rates of attraction within the domain. A strong attracting structure occurred in the northwestern part of the channel. As shown in figure 4-9.(c), a suggested experiment was to release drifters on each side of this structure, as noted by the pairs 1 and 3, 2 and 4 or 5 and 6. Advection of the structure and the pairs until 16:00 predicted that the drifters would align along the attracting structure.

As most of the LCS features were located within the channel, the plan for the experiment was to release several drifters around this area. As shown in figure 4-10 on page 157, 7



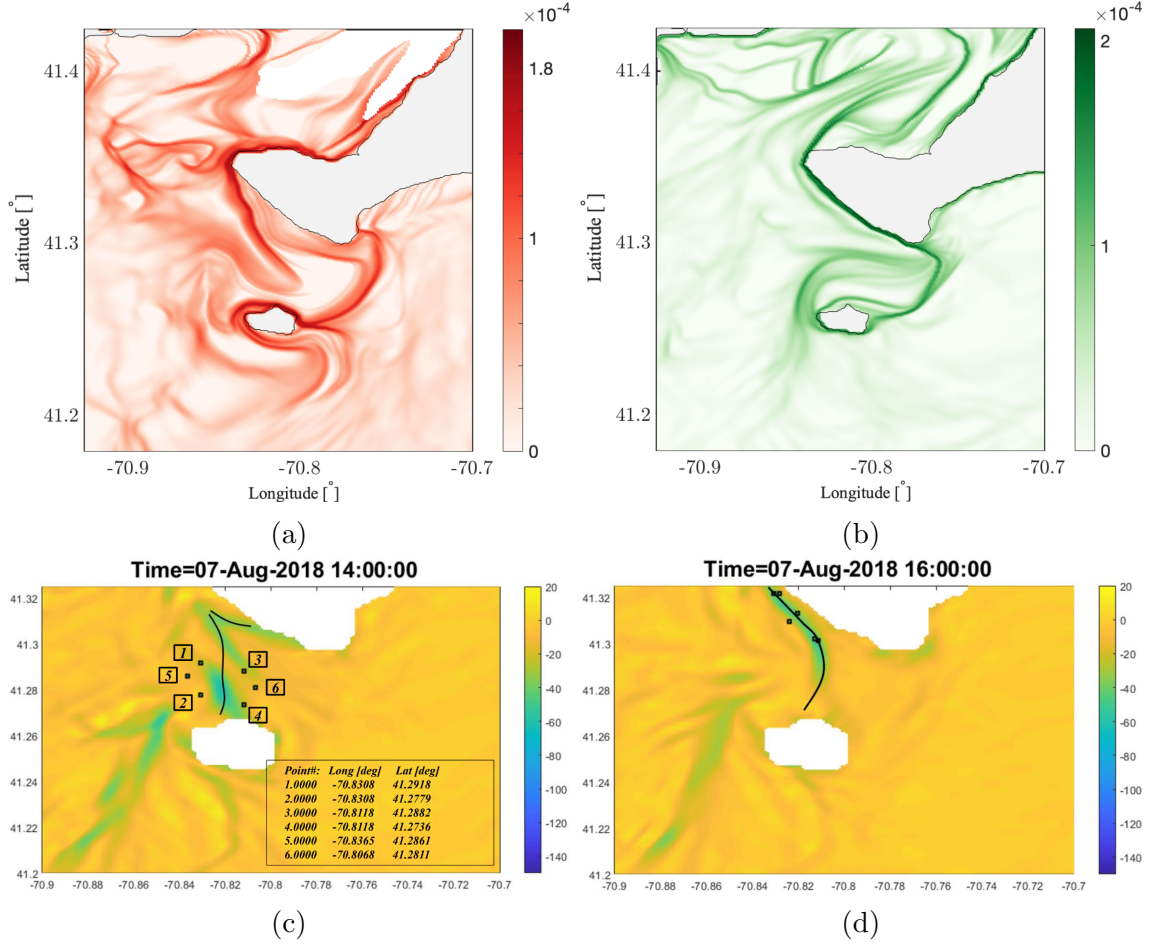


Figure 4-9: LCS predictions used for the 2018 No Man's Land experiment. (a) Forward-time 6-hour FTLE field computed by MSEAS for 16:00. (b) Backward-time 6-hour FTLE field computed by MSEAS for 16:00. (c) Attracting OECS computed for 14:00 by Mattia Serra. The numbers correspond to suggested release locations of pairs: 1 and 3, 2 and 4 or 5 and 6. (d) Locations of the attracting OECS and the proposed pairs at 16:00.

drifters were released within the channel alone: 1, 3 and 5 through 9. A group of 4 drifters were released east of the ridge and the remaining 7 drifters were spread south and west of No Man's Land. The expectation was that the first group of drifters would be advected northward while the eastern group would remain east. The LCS of interest being formed around 16:00, the goal was to start the deployment early so that all drifters would be in water by 16:00.

### 4.3.2 Experimental results

The first drifter was deployed in water at 14:01 and the last drifter was deployed at 16:00, meaning that all drifters were in water at the start of the time-window of interest. The experimental CODE drifter trajectories corresponding to the 6-hour release from 16:00 to 22:00 UTC on August 7, 2018 are shown in figure 4-10 and color-coded by time. The trajectories exhibited a behavior vastly different from the predictions. None of the drifters were advected northwest through the channel and toward the western tip of Martha’s Vineyard, as was predicted by FTLE and OECS. Drifters 1 and 3 were advected along the northern coast of No Man’s Land until drifter 1 beached while drifter 3 continued on an eastward trajectory. Drifters 2, 4 and 16 headed first southwest then east. All drifters were eventually advected eastward after 20:00.

### 4.3.3 Comparison of experimental and numerical trajectories

The comparison with numerical trajectories is shown in figure 4-11.(a) on page 158. Similarly to 2017, the numerical trajectories were seeded individually in TRACE using the MSEAS velocity output to use the exact same initial time and positions as the experimental trajectories. As was expected from the LCS predictions, some numerical trajectories ended up in the northwestern part of the channel. Some experimental drifters were advected in the opposite direction from their numerical counterpart, such as for the dark red and turquoise trajectories. The distances between experimental and numerical trajectories are shown in figure 4-11.(b).

A snapshot of the wind forecasts used in the model is shown for 18:00 in figure 4-12.(a) and indicates southwest winds, meaning the air flow was in the northeast direction. For comparison, measured wind data was retrieved for the Martha’s Vineyard Airport Station from The Weather Underground at [www.wunderground.com](http://www.wunderground.com) and from the National Centers for Environmental Information (NCEI) at [www.ncdc.noaa.gov/cdo-web/](http://www.ncdc.noaa.gov/cdo-web/). The Weather Underground’s time series for the observed wind data is plotted in figure 4-12.(b) for the day

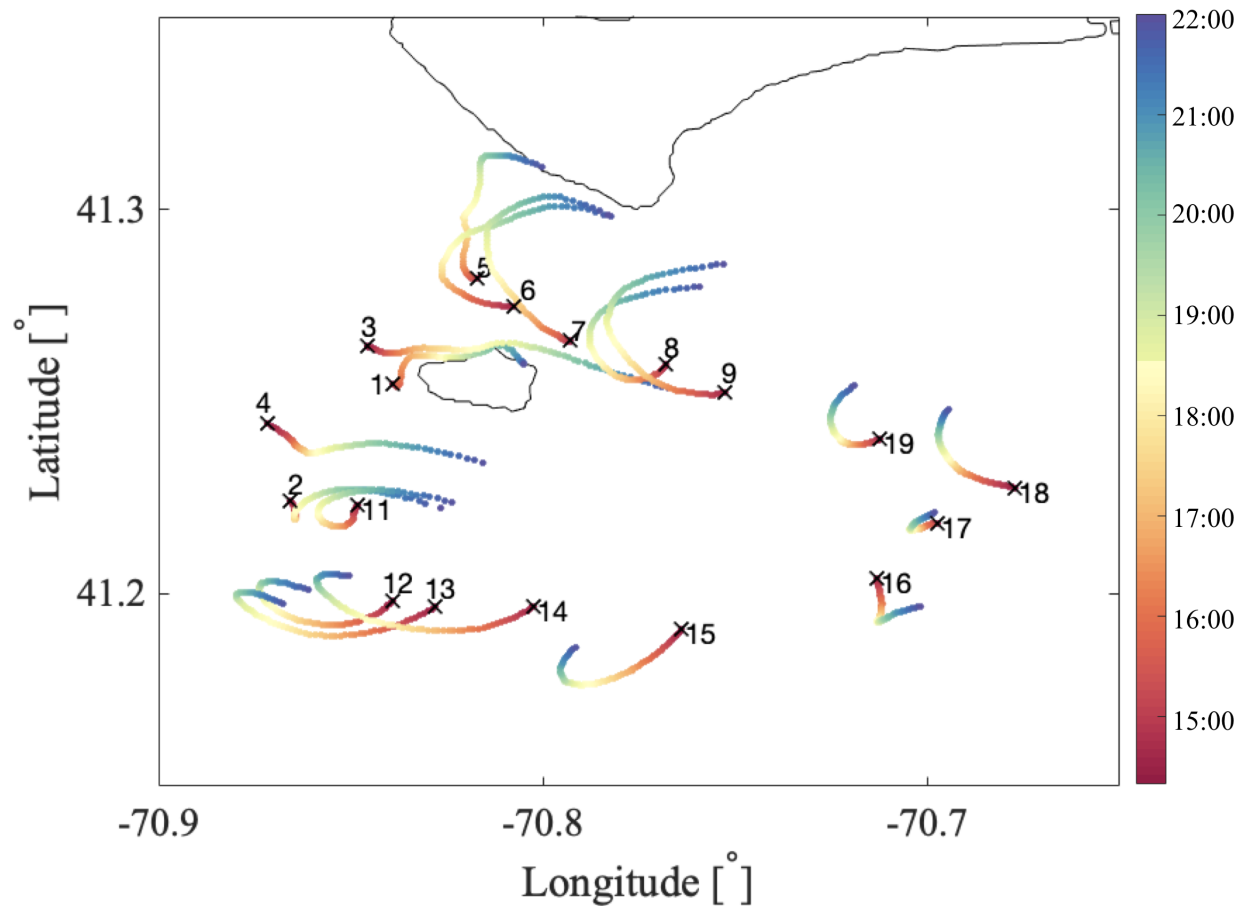


Figure 4-10: CODE drifter trajectories, color-coded by time, for the August 7, 2018 No Man's Land experiment until 22:00, 6 hours after the last drifter was released in water at 16:00. The drifters were numbered 1 through 19, with drifter 10 not included in this release.

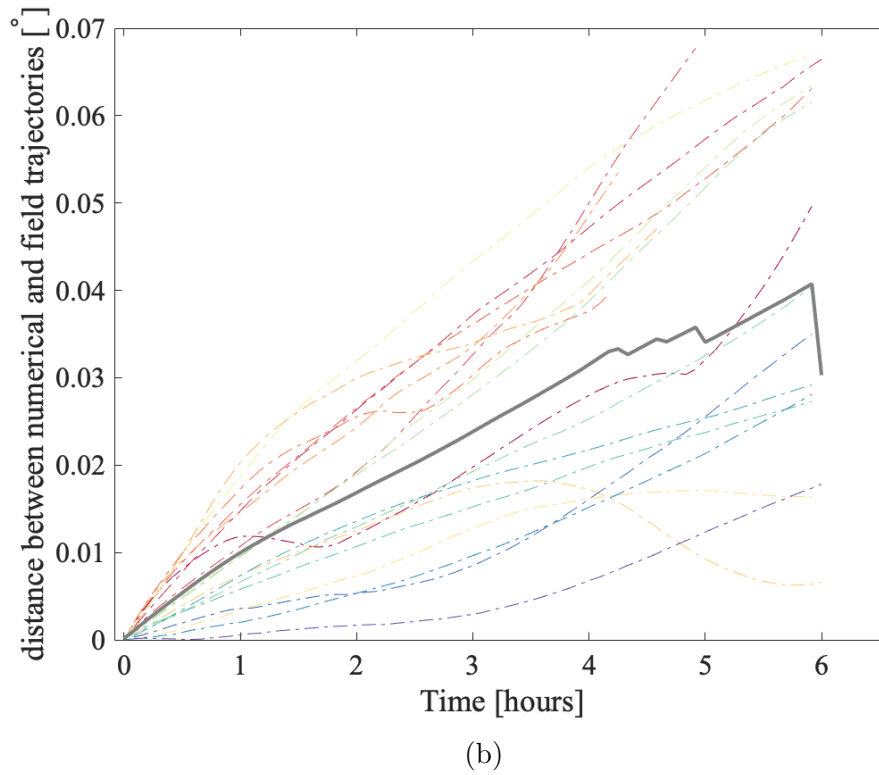
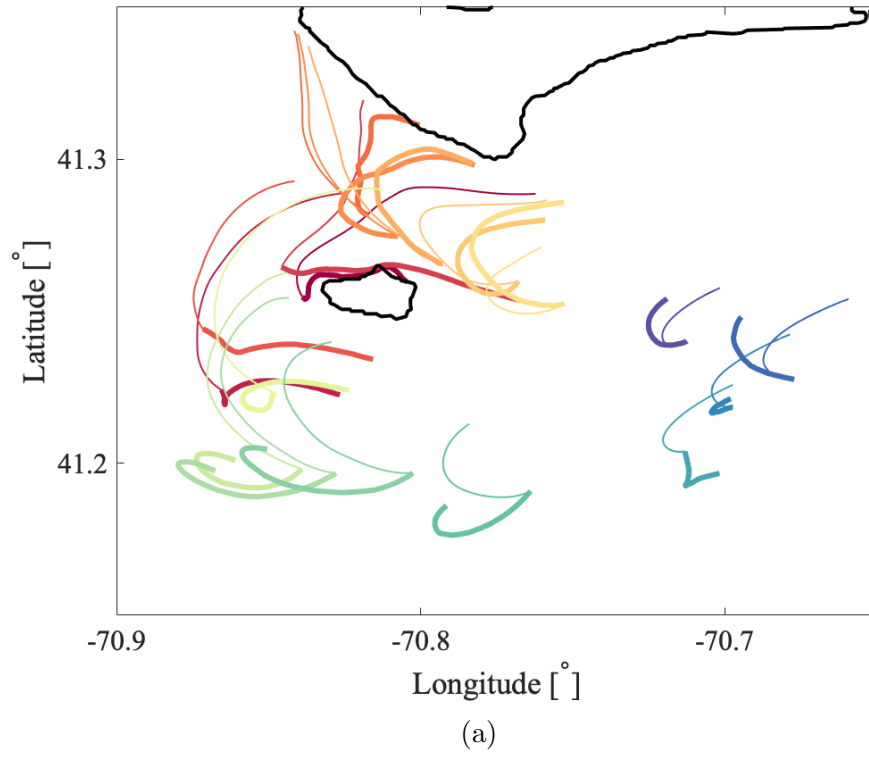
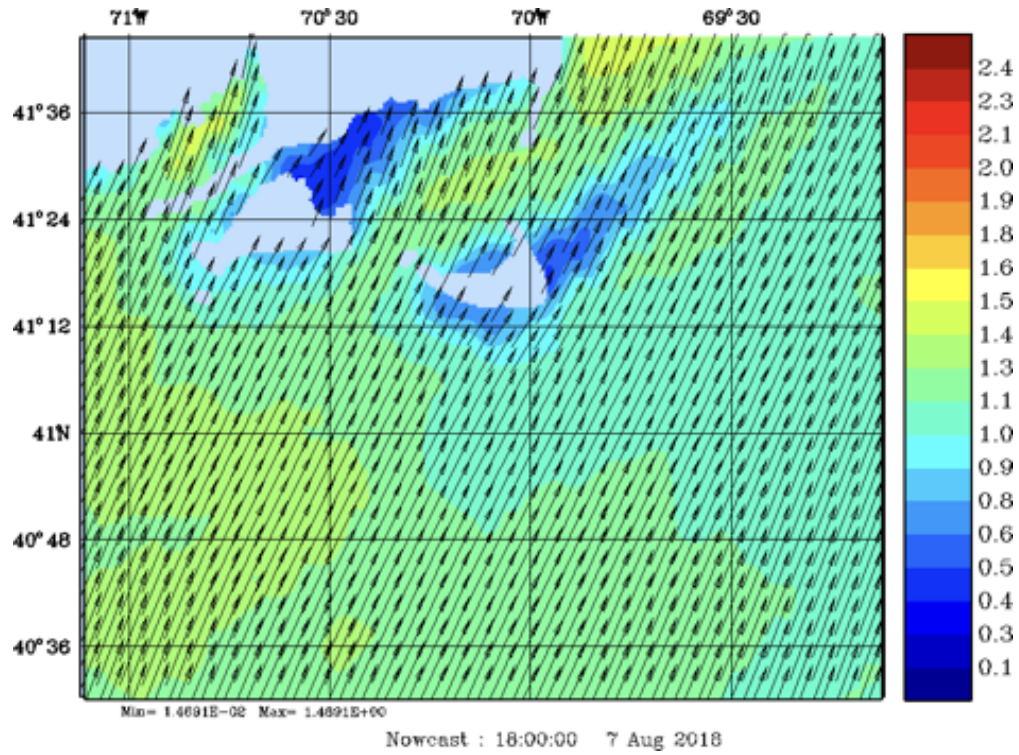
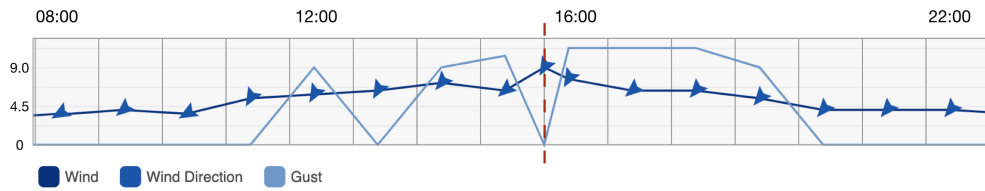


Figure 4-11: (a) CODE drifter trajectories (thick lines) and numerical trajectories (thin lines), individually seeded in TRACE from the MSEAS velocity fields. Each color correspond to a different drifter trajectory. (b) Pairwise distances between the experimental and numerical trajectories for each drifter. The colors match those in (a). The average distance is plotted with the thick grey line.



(a)



(b)

Figure 4-12: (a) Wind data used by MSEAS forecasts. (b) Measured wind time series from The Weather Underground for August 7, 2018 for the Martha's Vineyard Airport station. Speeds in m/s. Times in UTC. The start of the release is marked in the red dashed line and the end by the purple dashed line.

of August 7, 2018, showing that wind gusts occurred between 16:00 and 20:00 UTC. The time series from NCEI indicated wind gusts between 15:53 and 22:53 UTC. Both time series show that during the drifter release, the wind direction was mostly southwest, in opposite to the predictions, and that the region experienced high gusts of wind during most of the drifter release. The discrepancy between observed and numerical drifter trajectories, as well as in speeds along the tracks, is thus likely due to the wind forecasts that were issued and used for the model. The change in heading towards an eastern coast was observed for all drifters starting around 20:00 as well, once this gusts died down.

The HF radar system was not able to survey the surface velocity fields for this day; thus, the model could not be corrected with in-situ measurements. Calibrating and reconstructing Eulerian velocity fields from a sparse set of Lagrangian drifter trajectories is very challenging and, oftentimes, it yields inaccurate results. For the 6-hour experiment taking place from 16:00 to 22:00, comparisons between LCS predictions and experimental trajectories are thus limited by the differences being modeled and observed data. This discrepancy was due to gusts of wind: these events are, by definition, sudden random and therefore not representative of the flow field around Martha’s Vineyard.

#### **4.3.4 Comparison of experimental and numerical trajectories for the 04:00 to 10:00 time window on August 8**

The CODE drifters were in water for 20 hours, until 12:00 on August 15, for 17 out of 18 drifters. The full trajectories are plotted in figure 4-14.(a). Towards the end of these trajectories, several drifters were headed north towards the western tip of Martha’s Vineyard, as had been predicted for the 16:00 release. These trajectories spanning two tidal cycles, the 04:00 to 10:00 time window for August 15 corresponds to the same tidal forcing as the 16:00 to 22:00 time window for August 14. The tidal time series for the two time windows

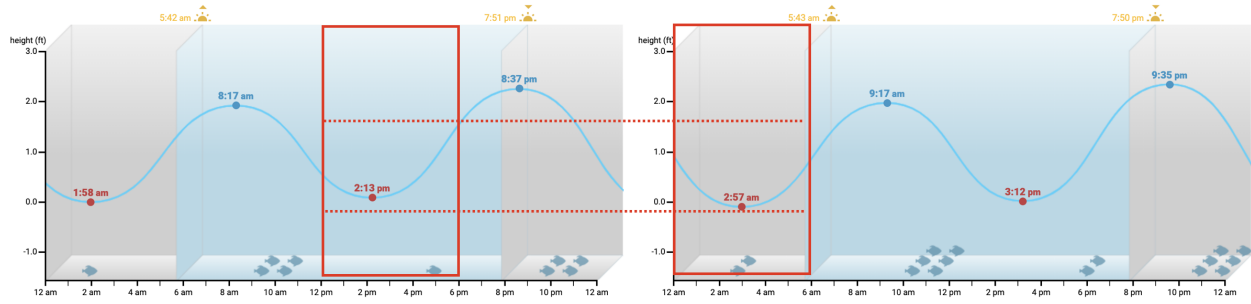


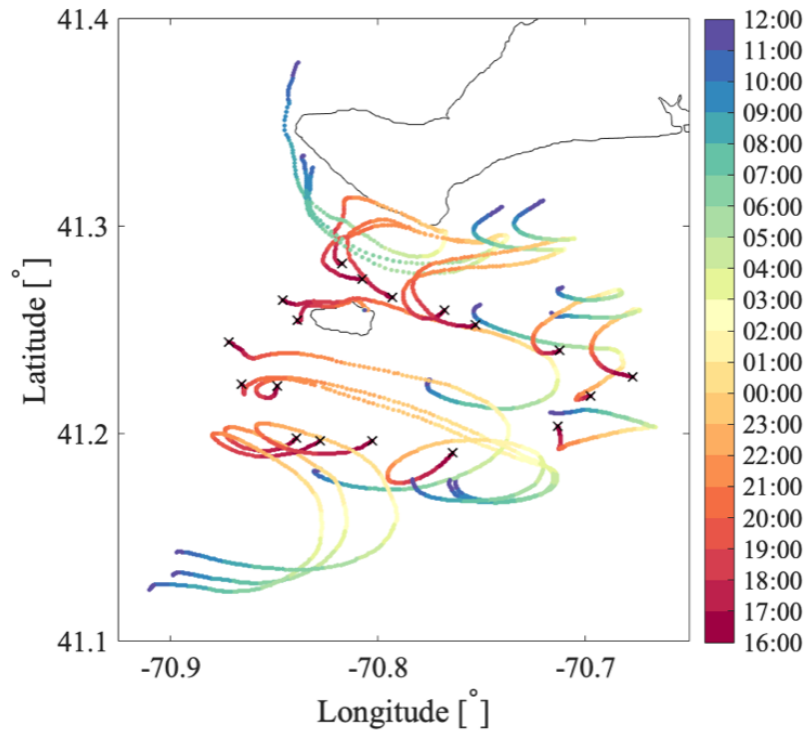
Figure 4-13: Tidal time series for August 7 and August 8, 2018. The 16:00 to 22:00 and the 04:00 to 10:00 time windows are indicated by the red rectangles. The dashed lines show the highest and lowest water levels over both time windows. Data from <https://tides4fishing.com/> for the north side of Muskeget island.

are shown in figure 4-13. In both cases, low tide occurred about half-way through the time window, explaining the west/east reversal of headings in the drifters trajectories in figure 4-11.(a) around 18:30 and in figure 4-14.(a) around 06:30.

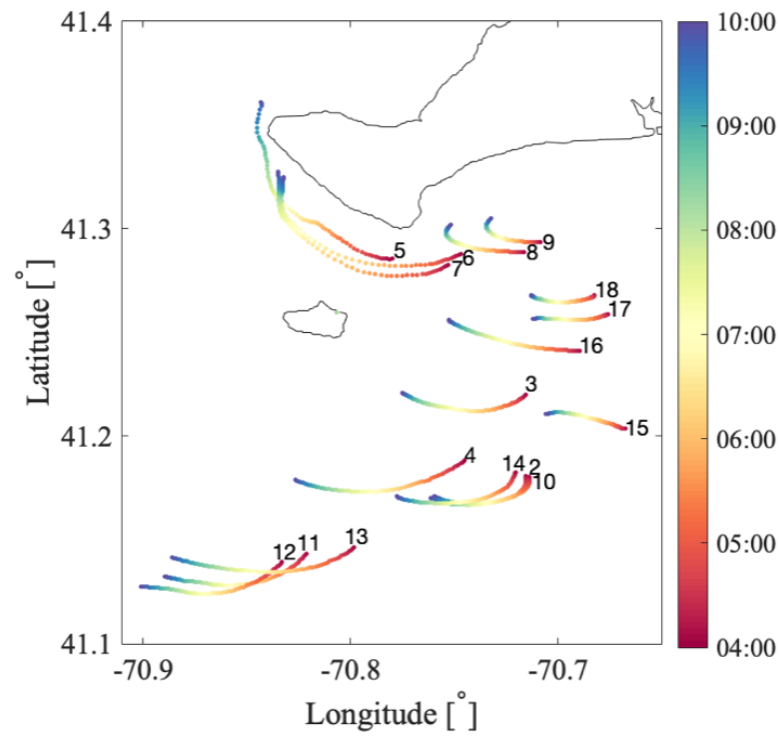
The LCS analyses are thus presented for the 04:00 to 10:00 time window. It corresponded to a similar point of the tidal cycle as the predicted experiment of 16:00, but the drifter trajectories, shown in figure 4-15 on page 163, exhibited the behavior that was expected during the planning of the experiment and the numerical trajectories are in much better agreement with the experimental data for this time window.

### 4.3.5 LCS analysis for the 04:00 to 10:00 time window on August 15

Similarly to the 2017 analysis, the FTLE, encounter volume, FCM and spectral clustering methods were used to analyze the model data from 04:00 to 10:00 on August 15, 2018. All methods were run with a 6-hour window of time integration. The results are plotted in figures 4-16–4-17 on pages 167–168. The left panel shows the methods at 04:00 and the right panel, at 10:00. The backward FTLE shows the results from 04:00 on August 15 to 22:00 on August 14 and from 10:00 to 04:00 on August 15. The encounter volume method was run on a larger domain, but on the 200-m grid as well. The clustering results are shown at



(a)



(b)

Figure 4-14: CODE drifter trajectories (a) until August 15 12:00, 20 hours after the last drifter deployment and (b) from 04:00 to 10:00 on August 15. The drifter positions are color-coded by time.



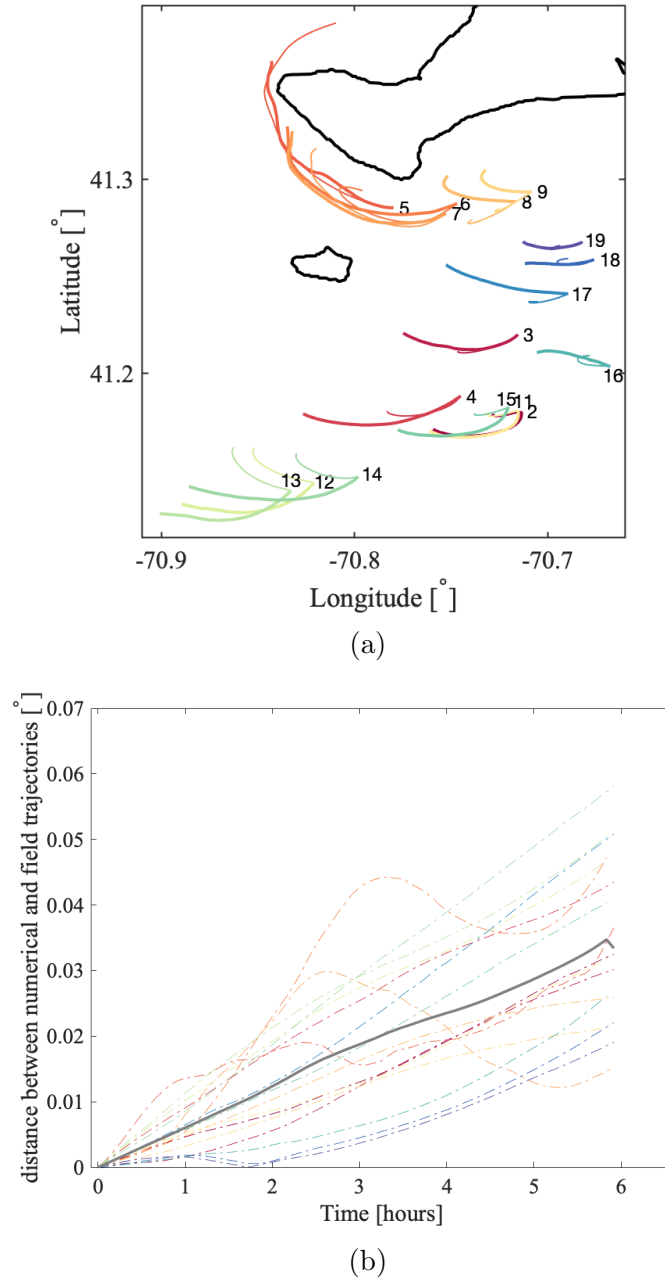


Figure 4-15: (a) CODE drifter trajectories (thick lines) and numerical trajectories (thin lines), individually seeded in TRACE from the MSEAS velocity fields, for 04:00 to 10:00 on August 15. (b) Pairwise distances between the experimental and numerical trajectories for each drifter. The average distance is plotted with the thick grey line.

the initial time (left) of 04:00 and the advected clusters are shown at the final time (right) of 10:00.

#### **4.3.5.1 Finite-Time Lyapunov Exponent (FTLE)**

The forward FTLE field (figure 4-16.a) showed high values of FTLE in the western part of the channel, with a ridge almost discernible between the eastern part of No Man's Land and the southern tip of Martha's Vineyard, resembling the picture from the 2017 FTLE analysis. High values were found west of Martha's Vineyard as well. One drifter, marked with a cross, was released west of the ridge, while two drifters, marked with diamonds, were released in close proximity of the ridge. These three drifters were advected northwest through the channel. The rest of the drifters, marked with asterisks, were released east or south of the FTLE ridge; none of these drifters crossed the channel. This difference in trajectory behaviors shows that the repelling forward FTLE ridge was a separating feature between the channel waters and the waters south of the Vineyard, suggesting a partition of the flow similar to that of 2017.

For the 10:00 to 04:00 time window, the backward FTLE field (figure 4-16.a') showed the highest rates of attraction along the coasts of Martha's Vineyard and No Man's Land, at the boundaries of the numerical domain. Three drifters were advected north through the channel, with one trajectory in particular closely following the ridge, validating the high rates of attraction.

#### **4.3.5.2 Encounter volume**

The encounter volume method was run with radii from 0.1 to 1 km. Convergence in the number of encountered trajectories was reached below a radius of 0.25 km. The 0.25 km results are thus the ones plotted in figure 4-16.b. The encounter volume method was run on a larger domain, which is shown in appendix C.2.2. The encounter volume values indicated high mixing potential inside the channel and south of No Man's Land. The high values

of encounter volume did not correspond to any FTLE ridge. For the 04:00 results, the high values of encounter volume were located inside the boundaries of the coherent spectral clusters shown in panel (d).

#### 4.3.5.3 Fuzzy C-Means (FCM)

The FCM analysis (figure 4-16.c) was run with  $K = 3$  clusters here as well to compare with the 2017 analysis. The drifters trajectories were superimposed and color-coded to correspond to the cluster in which they started and/or ended. The initial positions are marked with circles and the final positions with large crosses. The three drifters released within the purple cluster stayed within throughout the release and were advected towards the northwest. For the blue cluster, east of the channel, two drifters ended in the purple clusters and one drifter ended in the orange cluster; their final positions are shown with black crosses. All other drifters stayed within their corresponding cluster. The three drifters released within the orange cluster stayed within throughout the release.

#### 4.3.5.4 Parameter-free spectral clustering

The parameter-free spectral clustering protocol was applied to the 04:00 to 10:00 window of integration. The sweeps of  $r$  radii are shown in appendix C.2.1. Figure 4-16.(d) shows the spectral clustering results and coherence metrics detected for the peak gap ratio at  $r = 0.0325$ . Two coherent clusters were detected: one north of No Man's Land, with a coherence percentage of 73.4%, and one south of No Man's Land, with a much higher coherence value of 94.5%. The clusters boundaries did not correspond to the boundaries of the FCM clusters. Only one drifter was released within a cluster, the one marked in purple, which stayed within the cluster throughout the release. All other drifters are marked in black. Two drifters are marked in grey: they started in the incoherent background and ended at the boundary between the incoherent and coherent cluster; this is illustrated in figure 4-17 on page 168, which zooms in on these final positions. Advection of the spectral

clusters showed that the clusters shrunk in area as they headed north, especially the purple cluster that followed the coast of Martha’s Vineyard along the attracting backward FTLE ridge.

#### 4.3.6 Discussion

The 6-hour trajectories from 04:00 to 10:00 on August 15 were in fairly good agreement with the numerical trajectories. The 20-hour trajectories shown in figure 4-14.(a) exhibited three different transport behaviors: three drifters were advected north through the channel and west of Martha’s Vineyard; three drifters were advected westward, south of  $41.15^\circ$ ; and the rest of the drifters reversed direction twice, around the turn of the tide, heading eastward by the end of the release.

The behavior of the first group of drifters, drifters 5 through 7, can be explained with the FTLE fields, the encounter volume method and spectral clustering. They were released west of the forward FTLE ridge or on the ridge. They were then advected within the channel and northward towards the western tip of the Vineyard, following the backward, attracting ridge, with drifter 5 staying within the purple spectral cluster throughout the release. This purple cluster shrunk after 6 hours and coincided with high values of encounter volume, suggesting downwelling, which would be consistent with high values of backward FTLE.

The reversal of the tide was apparent in most of the trajectories that started east of the FTLE ridge. The change of direction did not seem to affect the trajectories that were below  $41.2^\circ\text{S}$ , which continued in a westward direction. This may be consistent with the mean southward drift mentioned in Rypina *et al.* [2014b], which is known to increase in strength with increasing distance from the Vineyard.

The three FCM clusters did not quite capture this north/south separation of trends. While the purple cluster was advected northward, several drifters from the blue cluster crossed the boundaries. This may be due to differences between numerical and observed data, but these discrepancies were not found in other methods.

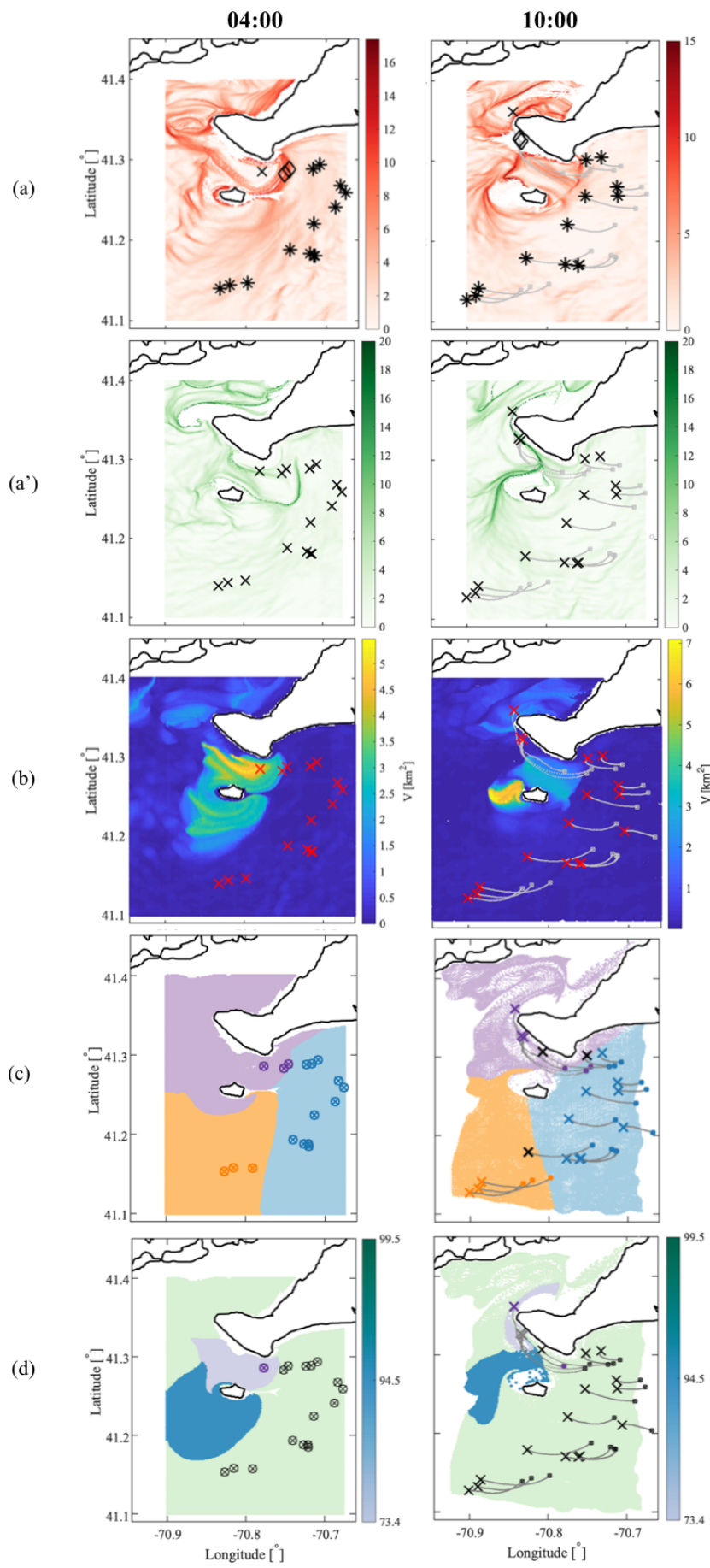


Figure 4-16: LCS and drifter positions for the 2018 experiment. (a) Forward and (a') backward FTLE, described in section 4.3.5.1. (b) Encounter volume (see 4.3.5.2). (c) Fuzzy C-Means clusters (see 4.3.5.3). (d-d') Spectral clusters with coherence metric (see 4.3.5.4).

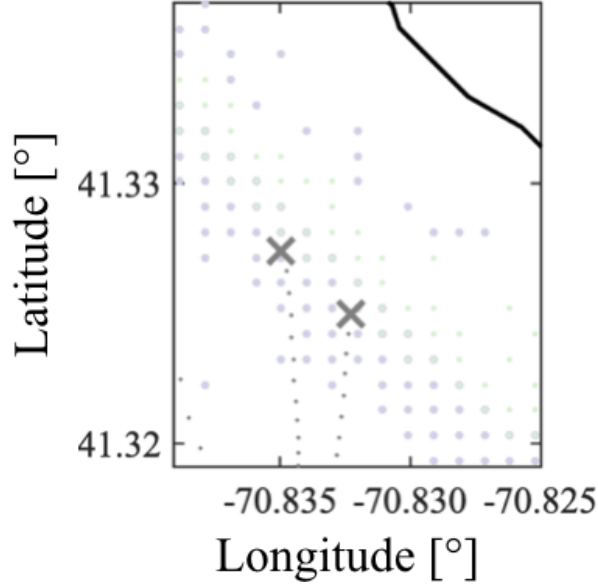


Figure 4-17: Zoom on the coherent spectral cluster at 10:00 from figure 4-16. The grey markers started in the incoherent cluster and ended at the boundary between the coherent and incoherent cluster.

## 4.4 Conclusions

The experiments demonstrated the feasibility of ocean flow studies at the submesoscale: in both the 2017 and 2018 experiments, the LCS analysis and the planning of the release happened about 12 hours prior to the departure of the vessel. Within the constraints of this last-minute planning, drifters were released around the predicted time and locations of the targeted structures. For both the 2017 and 2018 experiment, the HF radar system was planned to survey the domain of interest and measure surface velocity fields. This was not feasible due to equipment and logistical challenges. The comparison between drifter data and model data were nonetheless sufficient to conduct the LCS analyses.

Several LCS features were observed and/or modeled in both the 2017 and 2018 experiments. First, the presence of a repelling, forward FTLE ridge in the channel between No Man's Land and Martha's Vineyard suggests a frequently reoccurring barrier to transport separating the waters west of No Man's Land from the domain. Second, the backward FTLE

fields showed the highest rates of attraction along the coast of Martha’s Vineyard, with a ridge attracting material in its vicinity northward. Third, a spectral cluster north of No Man’s Land was present in the 2017 and the 2018 experiment: this cluster was advected northward along the attracting backward ridge in both cases.

The different methods offered a breadth of information about the flow transport. The FTLE, encounter volume and clustering methods were complimentary of each other, yet oftentimes were in agreement. The forward FTLE method proved itself to be a fast and accurate way to identify the different regions in the flow. The forward, repelling FTLE ridges partitioned the domain into an east and a west subdomains. Low regions of FTLE were indicative of low mixing regions and relatively low velocities, as seen by the short drifter trajectories. The backward FTLE method was equally accurate in identifying the key attracting structures and was validated by the trajectories of the drifters that were released in their vicinity.

The spectral clustering approach used the parameter-free protocol from chapter 2 and revealed the connectivity in the domain, for structures of different scales. The main spectral clustering result was the identification of a coherent cluster inside the channel that was advected northward toward the backward FTLE ridge. The shrinkage in cluster area and the relatively high values of encounter volume found within the cluster boundaries suggested that this cluster corresponded to a zone of convergence. This phenomenon was observed in 2017 and in 2018. In both cases, one drifter had been released within the cluster and stayed within its boundaries throughout the release, validating its coherence. A secondary result from the 2017 analysis was the identification of a region of high connectivity southeast of the channel, which agreed with the low FTLE values in the area, as well as the low mixing potential indicated by the encounter volume method.

The FCM clustering method allowed for a quick visualization of the partition of the domain from geometry, but the parameters had to be chosen in a subjective manner. The clusters generally did not agree with the spectral clusters. Moreover, the partition of the

2018 domain was invalidated by several drifter trajectories.

The encounter volume method provided useful information about the mixing potential within the domain. In some cases, the values of high mixing potential corresponded to the ridges of the FTLE; in other cases, the high values of encounter volume indicated convergence in the flow. In the latter case, the zones of high encounter volume were delimited by a spectral cluster, showing agreement with more than one LCS method. As a stand-alone approach, however, the encounter volume would not elucidate the different behaviors in drifter trajectories.

The 2018 analyses suffered from several set-backs. First, the wind gusts that occurred between 16:00 and 20:00 at the start of the 6-hour release on August 7 significantly affected the trajectories of the drifters. These gusts had not been predicted by the forecasts from the National Centers for Environmental Prediction (NCEP) used by the MSEAS model for the atmospheric forcing. The other forcing mechanism used in the circulation model being the tide, the time window starting at the next tidal cycle, at 04:00 on August 15, showed good agreement between the experimental and numerical trajectories. Second, the availability of the numerical domain was quite limited around the channel for the 16:00 to 22:00 time window of integration: for a lot of numerical particles seeded within the channel, the northeast flow direction seen in the numerical velocity fields resulted in advection toward the coast of Martha's Vineyard; these trajectories could therefore not be used in the 6-hour Lagrangian analysis. The coastal, backward FTLE ridges and the shrinkage of the spectral clusters along the boundaries may be due to material being advected onto shore and other coastal effects or simply due to computational limits. The 2018 field study was limited by differences between observed flows and model flows: indeed, the LCS analysis can only be as accurate as the velocity field used for processing.

The Martha's Vineyard case study revolved around a tidally-driven flow in a channel, in which a repelling FTLE ridge was formed at a certain point of the tidal cycle, around low tide. This is further illustrated in appendix C, which shows the evolution of the 6-hour FTLE



fields at 1-hour increments for 6 hours around low tide and the FTLE field at high tide. The ridge was first formed about an hour before low tide in the eastern part of the channel. It was progressively formed westward and disappeared after an hour past low tide. No such ridge was found in the channel at high tide. The influence of the tidal cycle on the formation of a pronounced FTLE ridge is similar to the Scott Reef case study from chapter 3, which also presented the LCS analysis of a tidally-driven flow. In both case studies, the FTLE analysis helped identified the key LCS delimiting regions with different transport behaviors and how this partition of the domain evolved in time, further showing the utility and applicability of the FTLE method for studying oceanic flows.



# Chapter 5

## Conclusion

This thesis investigated the present-day abilities to describe and understand the organization of flow transport at the ocean surface over the submesoscale range by advancing the theory and applicability of several Lagrangian Coherent Structures (LCS) methods. This chapter summarizes the advances that were presented in the previous chapters. The results are synthesized in section 5.1 and discussed in section 5.2. Potential avenues of future work are presented in 5.3.

### 5.1 Synthesis of results

The first result of this thesis was the development of an improved approach for spectral clustering, presented in chapter 2: a parameter-free protocol that detects coherent clusters and their surrounding incoherent background within the flow domain over a given time interval. With the increasing popularity of machine learning techniques and their applications to various disciplines, the field of dynamical system theory adapted a few algorithms for the detection of LCS. Of interest in this thesis were Fuzzy C-Means (FCM) and spectral clustering, two methods with results that heavily rely on the initial parameter settings. To avoid arbitrary choices, a parameter-free approach was introduced: it sweeps through parameters to find the peaks in gap ratio, which is the ratio of the maximum eigenvalue gap over the

range of eigenvalues. Based on the eigengap heuristics, the peaks in gap ratio correspond to a partition of the domain in which the coherent clusters have minimal inter-cluster similarity while having maximal intra-cluster similarity and where the incoherence of the background cluster is maximized.

The protocol included the computation of a noise-based metric to evaluate the relative coherence of the resulting clusters. The coherence was defined by the robustness to small, random noise and was computed by slightly and randomly perturbing the trajectories, then advecting them and counting the percentage of trajectories that remain within the boundaries of the clusters. The accuracy and the usefulness of the proposed protocol and of the coherence metric have been successfully demonstrated in three idealized model flows: the Bickley Jet, the Duffing oscillator and the asymmetric Duffing oscillator. The protocol performed well and was able to detect all the features of interest. Moreover, it detected the lobes as being part of the coherent clusters in both the symmetric and the asymmetric Duffing oscillator. These lobes corresponded to the filaments that were delimited by the ridges from the forward and backward FTLE fields (stable and unstable manifolds), proving that the parameter-free spectral clustering protocol can identify a domain partition with minimal inter-cluster similarity and with a highly incoherent background cluster.

The second main contribution of this thesis was the successful planning and execution of one of the first field studies targeting LCS in a predictive manner: the Scott Reef case study in chapter 3, which took place within a tidally-driven channel flow. Previous studies targeting LCS in oceanic flows had included Haza *et al.* [2010], who had done two similar field experiments with the Finite-Size Lyapunov Exponent (FSLE) using very low numbers of drifters, and Rypina *et al.* [2014b], who deployed dozens of drifters to investigate the complexity measure method. No drifter-based field study, however, had been conducted to target the more established Finite-Time Lyapunov Exponent (FTLE) method or the recently developed clustering methods. In the Scott Reef experiment, even sparse arrays of surface drifters enabled the experimental detection of a barrier to transport, which was

shown to correspond to a forward, repelling FTLE ridge, and of two groups of drifters, which corresponded to coherent spectral clusters. Further investigations of the physical processes underlying the observed drifter behavior and the spectral clustering results suggested that these clusters were caused by the downward velocities during the ebb at spring tide. Both the FTLE and spectral clustering highlighted the difference in flow transport at different times of the tidal cycle.

The third main contribution of this work was the comparison of a range of LCS methods in two field experiments: the No Man’s Land case studies of 2017 and 2018, which are presented in chapter 4 and also concern a tidally-driven channel flow. These field investigations covered a domain that was 30-kilometer-long, as opposed to the 8-kilometer-long array of drifters that was released at Scott Reef, while they both focused on transient features occurring over 6 hours only. Within these experimental constraints, the resulting field observations still captured the key dynamics of interest, through the release of Lagrangian drifters. For the 2017 and 2018 No Man’s Land experiments, the model velocity data was output the day prior to the experiment and the subsequent Lagrangian analyses were carried out within hours of the vessel’s departure. Nevertheless, the 2017 field trajectories validated the LCS predictions. The 2018 field experiments were challenged by unforeseen environmental conditions: wind gusts that significantly impacted the trajectories. Comparison between model data and experimental trajectories was still possible at a later time of the tidal cycle, showing the recurrence of key LCS with the tide.

## 5.2 Discussion

Predicting ocean flow transport is a scientific challenge because of the complexity of the underlying dynamics governing the flow. The No Man’s Land 2018 experiment was a prime example of how a disturbance in atmospheric forcing can yield vastly different results between drifter trajectories. The experimental investigations showed the utility and, sometimes, the

necessity to use several LCS methods to understand the organization of transport.

The case studies in chapters 3 and 4 proved the utility of using various LCS methods to gather different information about the flow of interest. The FTLE approach is very reliable as it always outputs a visualization of the separation rates between particles in the flow and clearly indicates where the high rates versus low rates of mixing are located. It is widely established and there exist several tools to make these computations very fast, such as the TRACE interface by Ameli & Shadden [2019] that was described in previous chapters. In the examples presented in thesis, the FTLE was sensitive to and revealing of three-dimensional effects within the domain, as well as the periodicity of the underlying dynamics. This was particularly obvious for the Scott Reef case study in which FTLE ridges were strongly dictated by the tidal cycle, with a sharply defined repelling structure being formed at high tide around spring tide. In the No Man’s Land case studies, a repelling, forward-time FTLE ridge was periodically reoccurring at low tide. In both cases, the ridges acted as barriers to transport between different regions of the flow. The different behaviors exhibited on different sides of the ridges were confirmed by drifter trajectories, especially in the case of Scott Reef, and/or by other LCS methods such as spectral or FCM clustering in the No Man’s Land studies. The main drawback of FTLE is that there is no established way to extract the structures of interest: the resulting scalar field is a diagnostic tool for visualization of separation rates and extracting the ridges involves the use of several parameters. These parameters include a threshold value to define ridges or criteria to extract continuous curves. In all cases, the FTLE predictions were experimentally validated, proving the robustness and applicability of the method.

The parameter-free spectral clustering approach proved to be particularly useful to extract coherent clusters of trajectories and understand connectivity within the flow. It helped identify and bound the areas of low mixing. Moreover, the sensitivity of the clusters to convergence in the flow helped detect where three-dimensional effects were the most important. Essentially, the spectral clustering approach is the K-Means algorithm, i.e. a special

case of FCM, applied to a distance function that has filtered the dataset by connectivity. This connectivity is also based on pairwise distances, but it has removed a large amount of insignificant entries. Moreover, from perturbation theory, the eigengap can determine the optimal number of clusters to be detected. This is the principal advantage over FCM, which is highly sensitive to the set of user parameters and for which the number of structures to detect has to be input. FCM will always yield a partition of the domain into geometry-based clusters, while spectral clustering may not yield useful clusters, but the absence of robust clusters has nonetheless some physical meaning: it indicates that the domain is homogeneous in connectivity. With improvements in computational efficiency, the parameter-free spectral clustering could become a more commonly used LCS method. The only structures that are extractable from FTLE are the ridges, which can be a tricky problem to implement [Senatore & Ross; Allshouse & Peacock, 2015a; Onu *et al.*, 2015], primarily because of a lack of objective criteria to determine the threshold values defining the ridges and because of the challenges of extracting continuous curves within the FTLE fields. One advantage that clustering algorithms offer over FTLE is that they allow the extraction of a different type of structures: coherent and incoherent clusters. Used along with the FTLE method, the parameter-free spectral clustering approach can provide an almost-complete description of flow transport.

The Lagrangian approach for oceanic flows ultimately depends on the accuracy of the available velocity fields. In spite of the utility of the approach, it is often limited by the quality of the model data. This is especially true for coastal areas, where the complex bathymetry, the boundary effects and the waves induce numerous uncertainties, even for state-of-the-art circulation models. In the Scott Reef and No Man’s Land channels, the domains were challenging to resolve numerically because of steep gradients in bathymetry and complicated coastlines. With recent improvements in satellite coverage accuracy and in high-frequency radar technology, the quality of the data assimilated in numerical models will likely lead to great enhancements in accuracy. Nevertheless, the FTLE analysis was able to predict

qualitatively correct features even with model uncertainties. This is primarily because the FTLE method outputs a global picture of the domain, whereas clustering methods outputs smaller structures within the domain; the difference in area between the different types of structures makes the clustering methods more challenging to validate experimentally and more sensitive to discrepancies between observed and model flows.

The introduction of the asymmetric Duffing oscillator chapter 2 stemmed from the intent to test the parameter-free spectral clustering protocol on a flow that contained coherent structures of different sizes. This flow was the outcome of a parameter space exploration that resulted in a regime with filamented gyres of different scales surrounded by a chaotic zone. This regime was very appealing for the study of LCS: while relatively simple, it exhibited differing gyres with interesting lobe dynamics. The detection of multi-scale features is arduous and challenging, in the area of LCS detection as well as in other disciplines. The parameter-free spectral clustering and the coherence metrics performed very well for this flow, which could become a benchmark test for methods from dynamical system theory.

This work presented two case studies of tidally-driven channel flows. In both the Scott Reef and Martha’s Vineyard studies, the organization of flow transport was strongly influenced by the tide. This was especially obvious for the repelling FTLE ridges that occurred sharply and periodically at high tide around spring tide in the Scott Reef channel and at low tide in the No Man’s Land channel. Preliminary FTLE and spectral clustering investigations of Scott Reef North indicated patterns that were similar to Scott Reef South: a repelling FTLE ridge and a coherent spectral cluster formed within the channel around high tide. In particular, the presence of separatrices, meaning repelling FTLE ridges acting as barriers to transport, may be due to the surface convergence associated with the tide, which separate the different bodies of water present around channel flows: typically, offshore waters versus *sheltered* waters, such as within a lagoon, an embayment or a channel. The evidence from the three field studies suggested that the presence of forward FTLE ridges, of coherent spectral clusters and of high values of encounter volume within the channels was connected to ocean



physics and to sea surface convergence. All studies being carried out in two-dimensions, this was seen from the numerical advection of particles and the divergence of the flows.

The main contributions of this thesis, beyond the advancement of the spectral clustering method by Hadjighasem *et al.* [2016], were the planning and execution of field experiments, which were assisted by many colleagues, as well as the comparative analyses of LCS methods. For chapters 3 and 4 of this thesis, the analyses were conducted on numerical model data that had been generated by colleagues. The collaborative efforts throughout this thesis show the importance of mixing different and complimentary expertises to apply theoretical methods to real-world oceanic flows, especially in highly interdisciplinary fields such as oceanography and ocean engineering.

### 5.3 Upcoming work

The next natural step is to extend the LCS analyses to other oceanic flows and, in particular, to other tidally-driven flows. The FTLE analysis of such systems could determine the prevalence of strong separatrices that are periodically occurring with the tidal cycle. The presence of transport barriers and of coherent clusters within three channel flows seemed to be due to surface convergence and it would be of high scientific interest to investigate whether this phenomenon is ubiquitous in the coastal ocean. Moreover, LCS may have the potential to highlight surface convergence or flow divergence in areas other than channel flows at the submesoscale, but further studies should be conducted to test the limits of the approach, especially in the open ocean. Coral reefs, especially atolls, have a well-defined bathymetry that shelters the ecosystem from offshore. The dynamics of the system of interest are thus confined within physical barriers: this, coupled with the periodic tidal or wave forcing that typically characterizes reef hydrodynamics, make coral reefs particularly amenable to LCS studies. The ability to predict robust transport barriers would greatly enhance the management of Marine Protected Areas (MPA) or the prevention of pollution events, including

contamination spills.

Another promising avenue of work is to use clustering methods, such as spectral clustering, to partition flows according to other physical properties, such as encounter volume. Chapter 1, for example, showed the encounter volume for the Bickley Jet. The areas of high and low encounter volume are easily detectable by eye, but the automatic extraction of structures is much more arduous. Through the spectral decomposition, the number of connected components is determined, meaning the number of clusters to retrieve is known. The areas of low mixing potential, such as the cores of vortices, could be extractable in a systematic manner. Sonnewald *et al.* [2019] used K-Means clustering to investigate the geography of ocean basins based on their barotropic vorticity. The number of clusters to detect was chosen from the convergence of the Akaike and the Bayesian information criteria, two criteria that estimate the quality of statistical models. A similar study could be undertaken to determine the number of connected components from spectral decomposition instead. Beyond vortices and Lagrangian Coherent structures at the submesoscale, the extraction of meaningful connected components of the flow has multiple applications in oceanography and geophysics.

# Appendix A

## Parameter-free spectral clustering protocol with the maximum distance function

The parameter-free spectral clustering protocol proposed in chapter 2 uses the average distance function. The approach was applied to idealized model flows, the Bickley Jet and the Duffing oscillator. To show the rationale behind the choice of the average distance function, the comparison with the maximum distance function is included here.

### A.1 Bickley Jet

Following the proposed parameter-free spectral clustering approach, the first step is to determine the gap ratio dependence on the radius  $r$ . The sweep of  $r$  values is shown in figure A-1a for an initial offset coefficient of  $10^{10}$ . The maximum distance function yielded slightly smaller gap ratios than the average distance function. Here, the peaks in gap ratios were found for  $r = [0.5, 3.0]$ . The convergence of parameters for offset coefficients  $10^n$  is verified in all cases; an example is shown for the peaks from the maximum distance function. In all cases, the spectral gap occurs after  $k = 6$  eigenvalues and the eigengap converges at offset

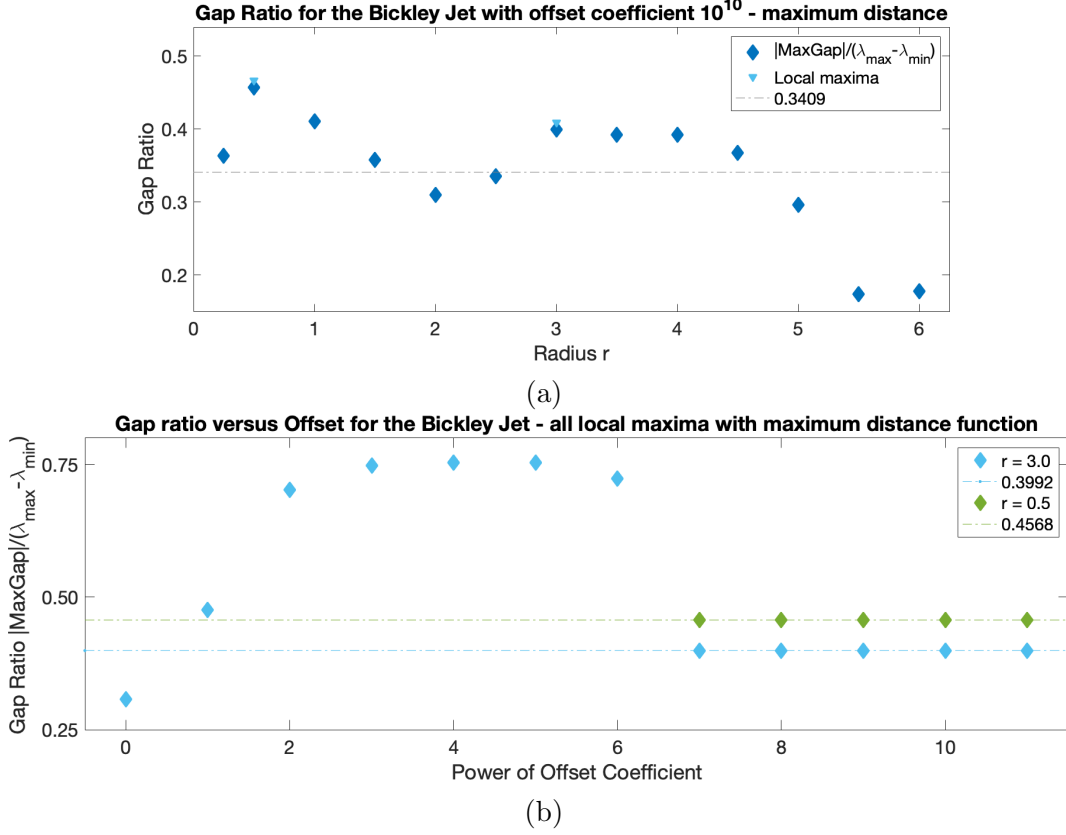


Figure A-1: Steps 1 to 3 of the spectral clustering protocol for the Bickley jet with the maximum distance function. (a) Gap ratio as a function of  $r$ . (b) Sweep of offset coefficients  $10^n$  for the gap ratio peaks  $r = 0.5$  and  $r = 3.0$  from (a).

value of  $w_{i,i} = 10^7$ .

To demonstrate the clustering results and the coherence metrics method, all local maxima in gap ratios were investigated. The coherence metrics was computed for all peaks, as shown in figure A-2 on page 183. In all cases, the protocol successfully detected the 6 recirculation vortices around the vicinity of the closed orbits in the Poincaré map. The sizes of the clusters varied with the radius  $r$  of sparsification, with very small clusters detected for  $r = 0.5$  and larger clusters for  $r = 3.0$ . The clusters detected with the maximum function contained little outside filamentation, as expected. All gyres had coherence percentages of similar order of magnitude, as was the case for the average distance function, but the key difference is the coherence value of the incoherent background: here, the background cluster had a coherence value of 100%. For the maximum gap ratio that occurred at  $r = 2.0$  with the average distance

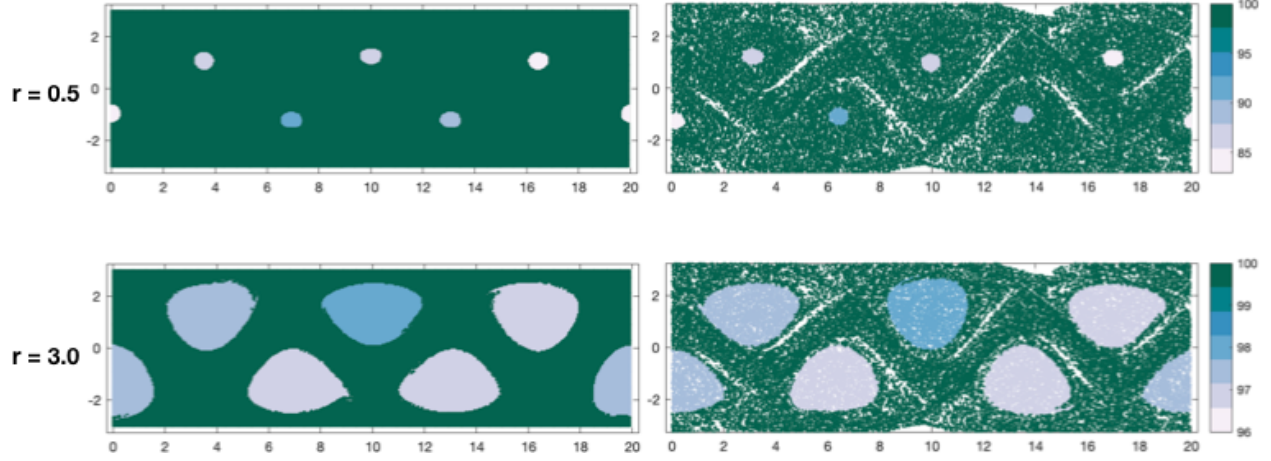


Figure A-2: Step 3 of the spectral clustering protocol for the Bickley jet with the maximum distance function: coherence metrics for the gap ratio peaks (Top)  $r = 0.5$  and (Bottom)  $r = 3.0$ . The (Left) initial and (Right) final positions are plotted.

function, the vortices had a coherence metrics higher than the incoherence background.

## A.2 Duffing oscillator

The second benchmark flow against which the parameter-free spectral clustering protocol was tested is the Duffing oscillator. The first step, the sweep in  $r$ , revealed two peaks in gap ratios at  $r = [0.75, 1.5]$  for the maximum distance function. The sweeps of parameters are shown in figure A-3 on page 184, followed by the second step of the protocol, the convergence verifications for the offset. For all cases, convergence in gap ratios was attained after offset coefficient  $10^7$ . The convergence plot includes the peaks for average distance function, which yielded larger gap ratios on average. The final step of the protocol, the coherence metrics results for each cluster, is shown in figure A-4 on page 184.

With the maximum distance function, two clusters were detected in each case, one around each gyre. The gap ratios corresponded, in increasing order, to  $r = 0.75$  and  $r = 1.5$ . The two peaks resulted in clusters that are roughly delimited by closed orbits in the Poincaré section. The closed orbits are isolated from the rest of the flow. For the larger gap ratio at  $r = 1.5$ , the incoherent background has a lower coherence value than the clusters. The

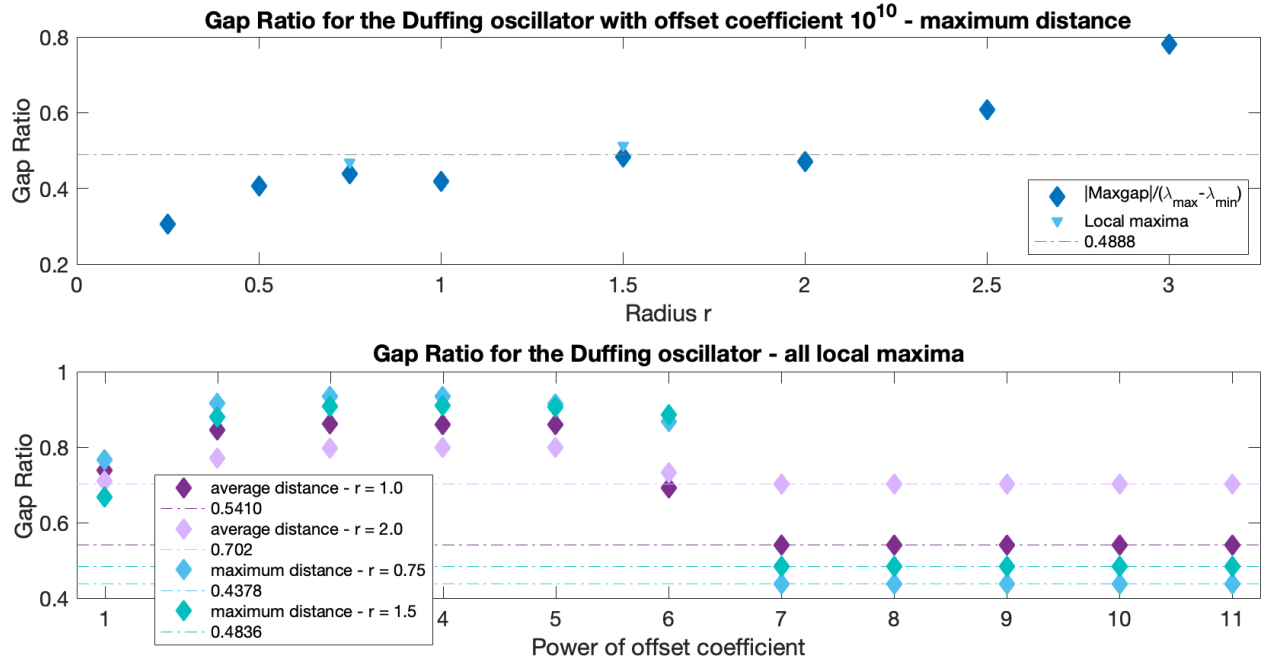


Figure A-3: Steps 1 and 2 of the spectral clustering protocol for the Duffing oscillator. (Top) Sweep of  $r$  parameters with offset coefficient  $10^{10}$  and the average distance function. (Middle) Same for the maximum distance function. (Bottom) Sweep of offset coefficients  $10^n$  for the gap ratio peaks.

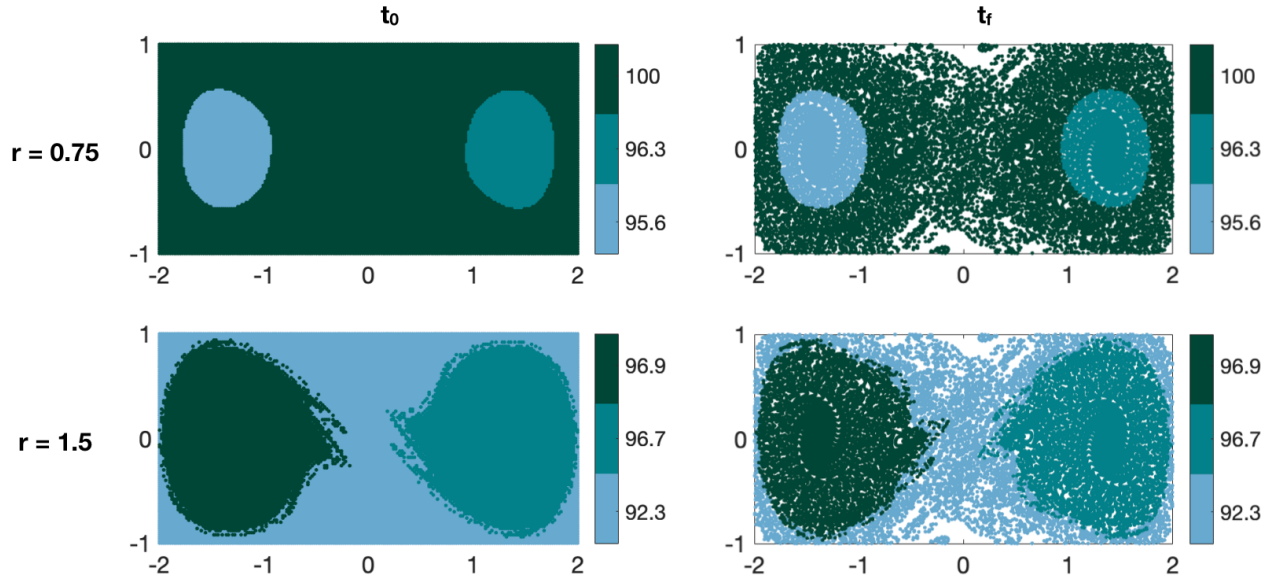


Figure A-4: Spectral clustering results for the Duffing oscillator for the maximum distance function: coherence metrics under noise. The results are shown for the two peaks at (Top)  $r = 0.75$  and (Bottom)  $r = 1.5$ . Initial and final positions are shown in each case, in the left and right panels, respectively.

clusters are not delimited by the FTLE ridges, however, as was the case for the clusters from the average distance function that included the lobes and had higher gap ratios. This suggests that focusing on the eigengap argument is a good way to determine the optimal spectral clustering partition of the domain.





# Appendix B

## FTLE ridge calculations for Scott Reef

The extraction of ridges of a Finite-Time Lyapunov Exponent (FTLE) field is a notoriously arduous problem in dynamical system theory [Allshouse & Peacock, 2015a]. FTLE ridges refer to the locally maximum values of FTLE that are connected along a curve [Shadden *et al.*, 2005], which correspond to structures with the strongest separation rates for forward time advection. In backward time, they correspond to the strongest rates of repulsion. FTLE fields and/or ridges are computed for the chosen window of time  $t_0$  to  $t_1$ . When FTLE fields are computed sequentially over a sliding integration window  $[t_0 + dt; t_1 + dt]$ , the FTLE ridges are not time-evolving structures, as they are sequentially recomputed over each interval. Each interval, indeed, represents a different finite-time dynamical system. The resulting sliding-window FTLE ridges are nonetheless sometimes considered to be Lagrangian Coherent Structures, but within certain limitations [Shadden *et al.*, 2005; Haller, 2015; Hadjighasem *et al.*, 2017]. The distinction between advected ridges from  $t_0$  to  $t_1$  and sliding-window FTLE ridges is important to make.

The Scott Reef case study in chapter 3 was planned and executed based on predicted forward FTLE ridge at high tide around spring tide, which occurred on October 3, 2016.

The primarily targeted feature that the FTLE ridge. At high tide around spring tide, a ridge is formed periodically south of the channel. Throughout the ebb, from 06:00:00 to 12:00:00 UTC, ridges are formed sequentially and progressively north of the channel. The deployment of drifters for the field experiment lasted almost 1.5 hour out of a 6-hour ebb tide, so the release started a bit after high tide, at 07:30:00, time at which all drifters had been deployed in water. The 6-hour window for the release thus ended at 13:30:00.

To investigate the sequential formation of forward- and backward-time FTLE ridges throughout the drifter release, FTLE fields were computed on sliding 6-hour windows from 07:30:00 to 13:30:00. In parallel, the extracted FTLE ridges from 07:30:00 were advected numerical until 13:30:00, 6 hours later. The advected ridges were superimposed to the recomputed fields for comparison, shown for the backward FTLE ridge in figures A-2-B-2 on pages 189-190 and for the forward FTLE ridge in figures B-3-B-6 on pages 191-194.

For ease and speed of calculations of both tracer trajectories and FTLE, the model data was uploaded in NetCDF format to a web-based gateway by Ameli & Shadden [2019], TRACE, which stands for Trajectory Reconstruction and Analysis for Coherent Structure Evaluation. TRACE uses parallel processing for the Lagrangian computations and can handle and process very large datasets that are hosted remotely. More information can be found at <http://transport.me.berkeley.edu/restore>.

In backward time, the superimposition of the advected ridge and the re-computed ridges matched almost perfectly. As pictured in figures A-2 and B-2, the ridges from the advection and from the sliding-window integration are not discernible from each other. In forward time, however, the advected ridge from 07:30:00 followed the re-computed ridge for a few hours, up to 09:00:00, after which the shape of the advected ridge stopped being a continuous curve and broke into a few pieces, as shown in figure B-4. Starting around 10:30, the advected ridge starts shaping into the recomputed backward-time ridge, before eventually transforming into the backward-time ridge, as shown in figures B-5 and B-6.

The forward-time ridge marking the highest values of separation between particles, any

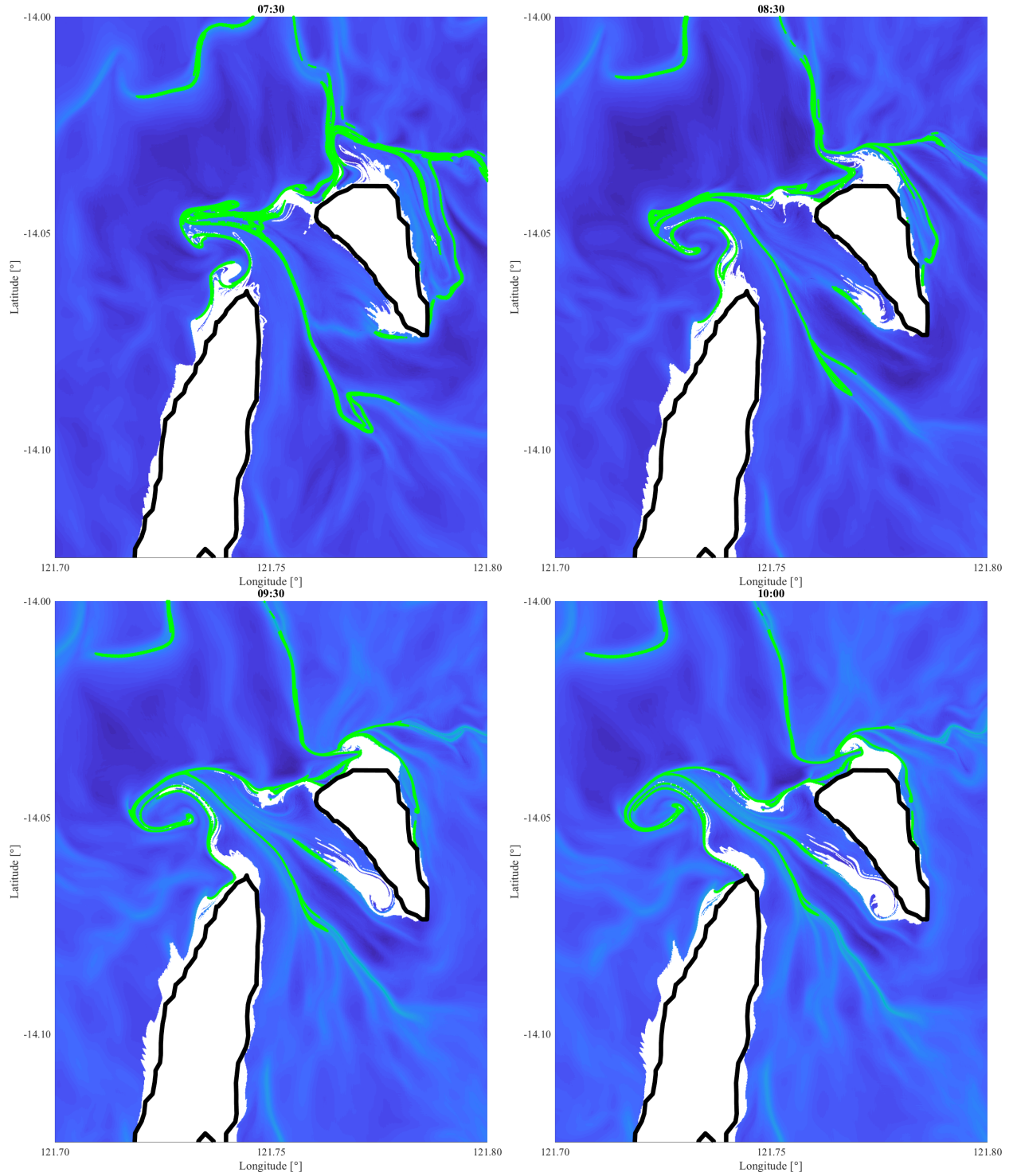


Figure B-1: Comparison between backward-time FTLE ridge advected from 07:30 and computed backward-time FTLE fields on a sliding window covering the Scott Reef field experiment (Part 1 of 2). The backward-time FTLE ridge (green) advected from 07:30:00 is superimposed to the 6-hour backward-time FTLE field recomputed at each time instant. The ridges from the advection and from the sliding-window integration are not discernible from each other.

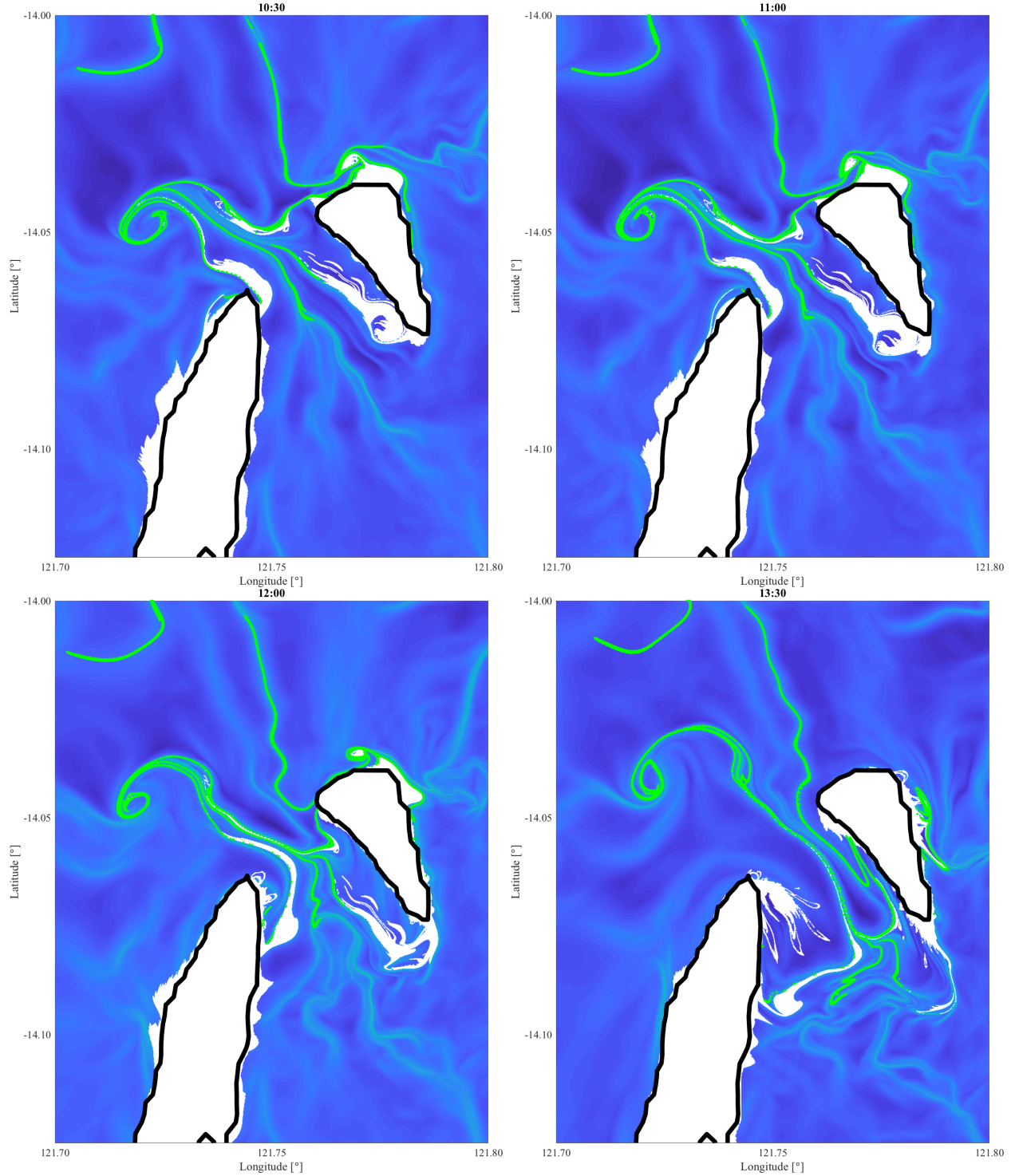


Figure B-2: Comparison between backward-time FTLE ridge advected from 07:30 and computed backward-time FTLE fields on a sliding window covering the Scott Reef field experiment (Part 2 of 2). The backward-time FTLE ridge (green) advected from 07:30:00 is superimposed to the 6-hour backward-time FTLE field recomputed at each time instant. The ridges from the advection and from the sliding-window integration are not discernible from each other.



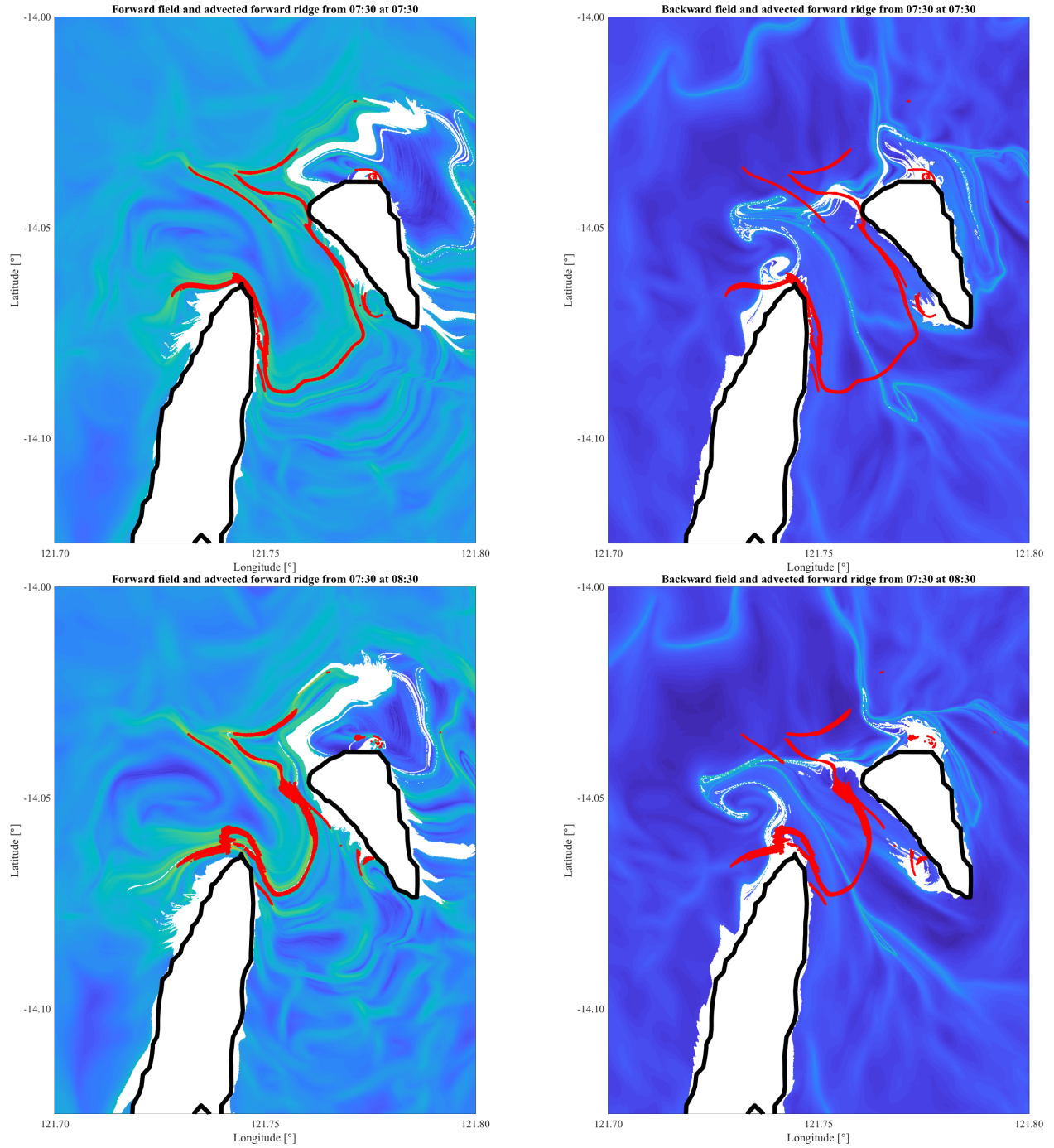


Figure B-3: Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 1 of 4). The forward-time FTLE ridge (red) advected from 07:30:00 is superimposed to the 6-hour forward- and backward-time FTLE fields recomputed at each time instant. Up to 08:30, the ridges from the advection and from the forward-time sliding-window integration are not discernible from each other.

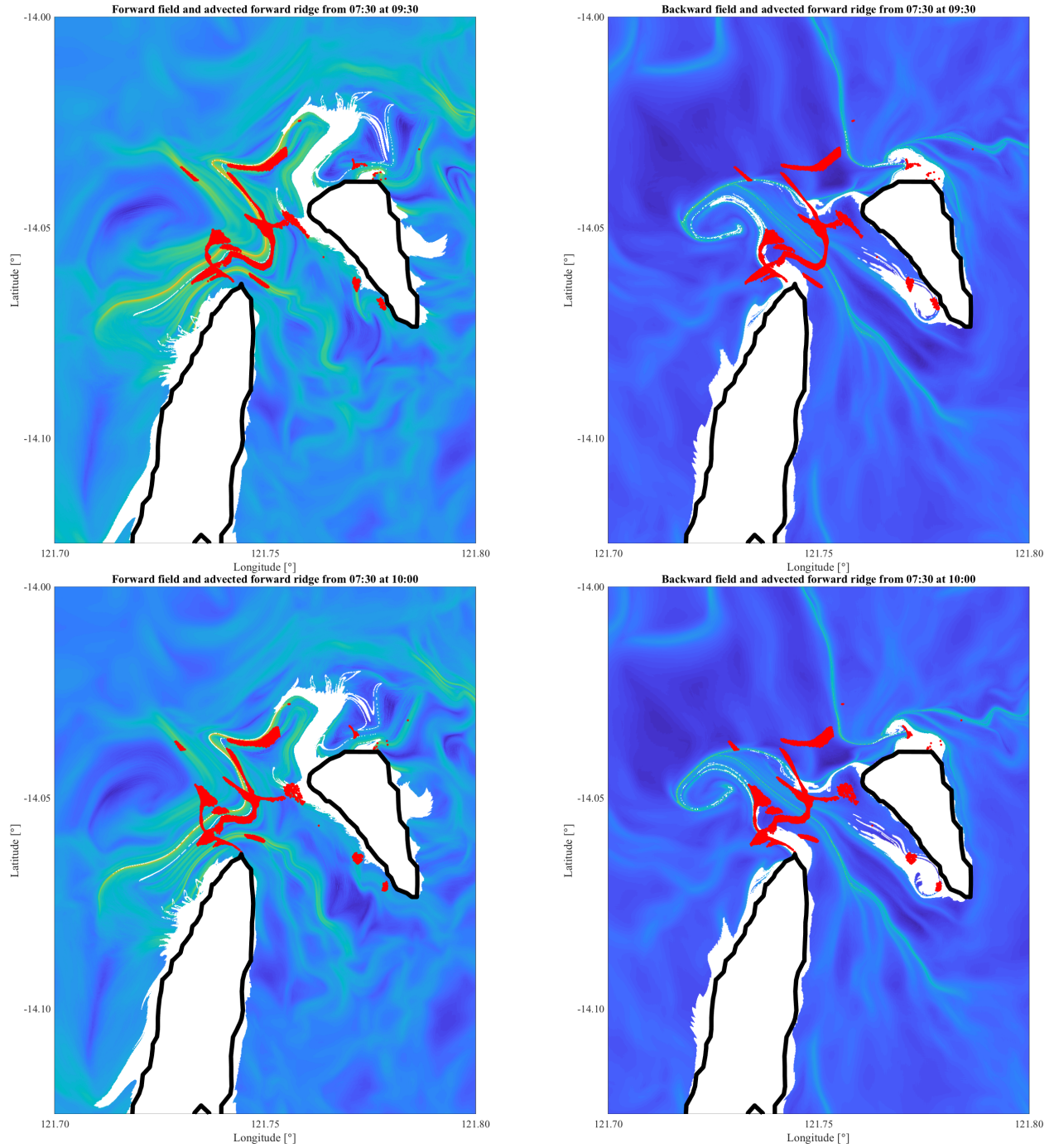


Figure B-4: Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 2 of 4). The forward-time FTLE ridge (red) advected from 07:30:00 is superimposed to the 6-hour forward- and backward-time FTLE fields recomputed at each time instant. The shape of the advected ridge starts to break and to deviate from the ridge of the forward-time sliding-window integration.

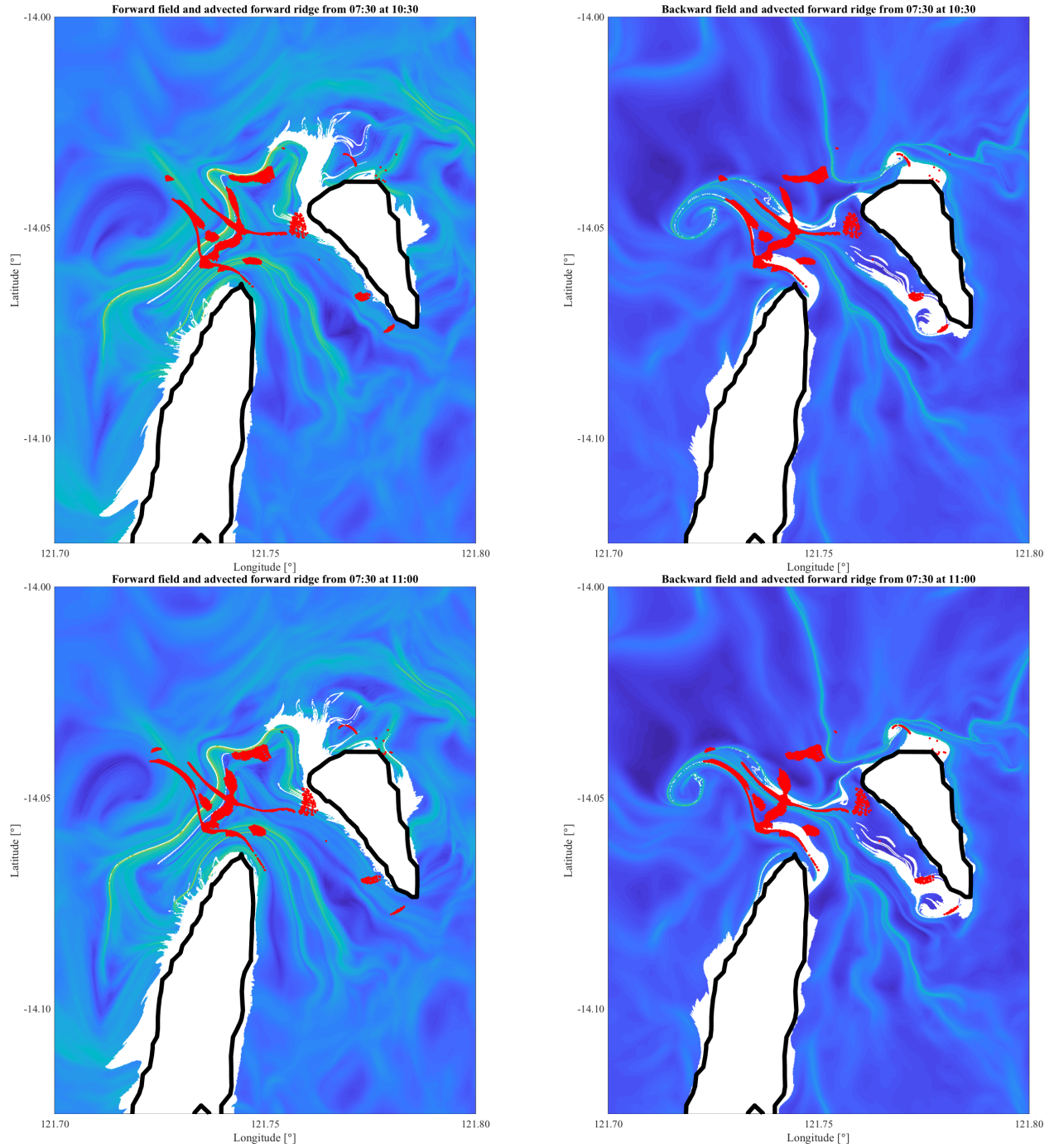


Figure B-5: Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 3 of 4). The forward-time FTLE ridge (red) advected from 07:30:00 is superimposed to the 6-hour forward- and backward-time FTLE fields recomputed at each time instant. One piece of the advected ridge still corresponds to the ridge of the forward-time sliding-window integration, but other pieces start to follow the shape of the ridge of the backward-time FTLE field.



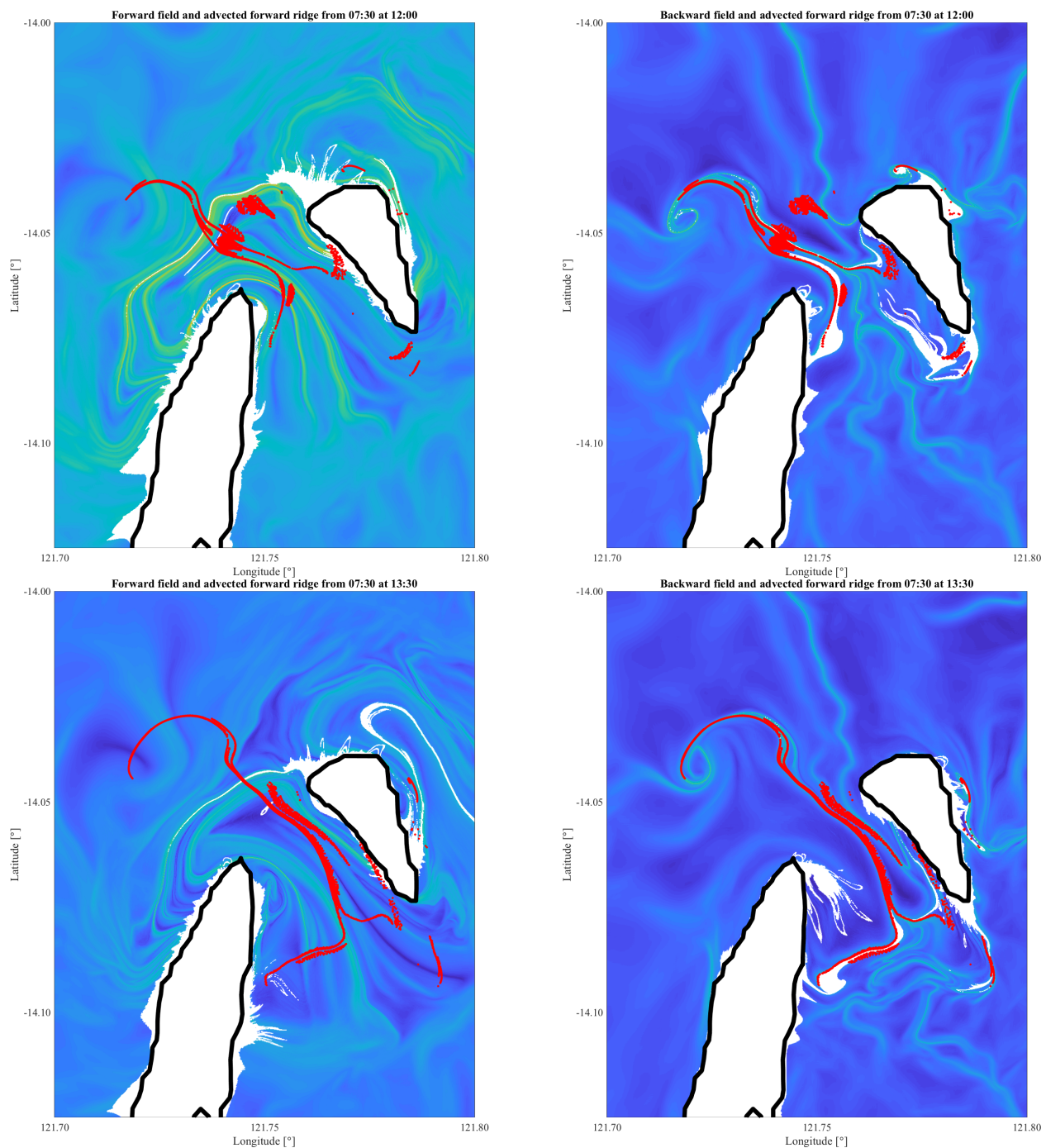


Figure B-6: Comparison between forward-time FTLE ridge advected from 07:30 and computed FTLE fields on a sliding window covering the Scott Reef field experiment (Part 4 of 4). The forward-time FTLE ridge (red) advected from 07:30:00 is superimposed to the 6-hour forward- and backward-time FTLE fields recomputed at each time instant. By the end of the 6-hour release, the advected ridge has now almost totally espoused the shape of the ridge of the backward-time FTLE field.



small deviation from the ridge's location due to numerical advection can result in the tracer particles separating from each other at high rates, which explains why the ridge start to break down after a few hours. On the opposite, the backward-time ridge marks the highest value of attraction, meaning that any tracer particle in the vicinity get attracted to the ridge. As both forward- and backward-time ridges are located in the channel, they are in proximity of each other. Thus, the tracer particles marking the forward-time ridge get separated at exponential rates, then attracted to the backward-time ridge at exponential rates. This shows that the forward-time ridge is a strongly repelling structure and the backward-time ridge, a strongly attracting structure.

The ridges computed on sliding integration windows do not form LCS *per se* as the overall integration time is not continuous. The comparison between advected ridges and recomputed ridges, nevertheless, shows that in backward-time, the sliding-windows ridges correspond to the advected ridge. For the forward-time ridge, the strong repulsion rates suggest that a repelling structure is present throughout the release. Some pieces of the advected ridge remain superimposed to the recalculated ridge for several hours, suggesting that this ridge is a persistent feature. For illustration purposes, the recomputed ridges are shown in figures 13 and 14 of chapter 3.



# Appendix C

## Details on Lagrangian computations for the Martha’s Vineyard case study

This appendix provides additional information about the results of some of the Lagrangian Coherent Structures methods utilized in Chapter 4 for the No Man’s Land experiment. Section C.1 concerns the results from the 2017 experiment and presents the Finite-Time Lyapunov Exponent (FTLE) dependence on the tidal cycle, the sweep of radii for the parameter-free spectral clustering methods and the results of the encounter volume method for the full domain. Section C.2 concerns the results from the 2018 experiment and presents the sweep of radii for the parameter-free spectral clustering methods and the results of the encounter volume method for the full domain.

### C.1 2017 No Man’s Land Experiment

#### C.1.1 Finite-Time Lyapunov Exponent (FTLE)

The 6-hour FTLE analysis was conducted at 1-hour increments from 12:00 to 17:00 on August 14, 2017 to show the evolution of the FTLE fields around the drifter release. On this day, low tide occurred at 15:02 UTC, as shown in figure C-1 on page 198, roughly half-way

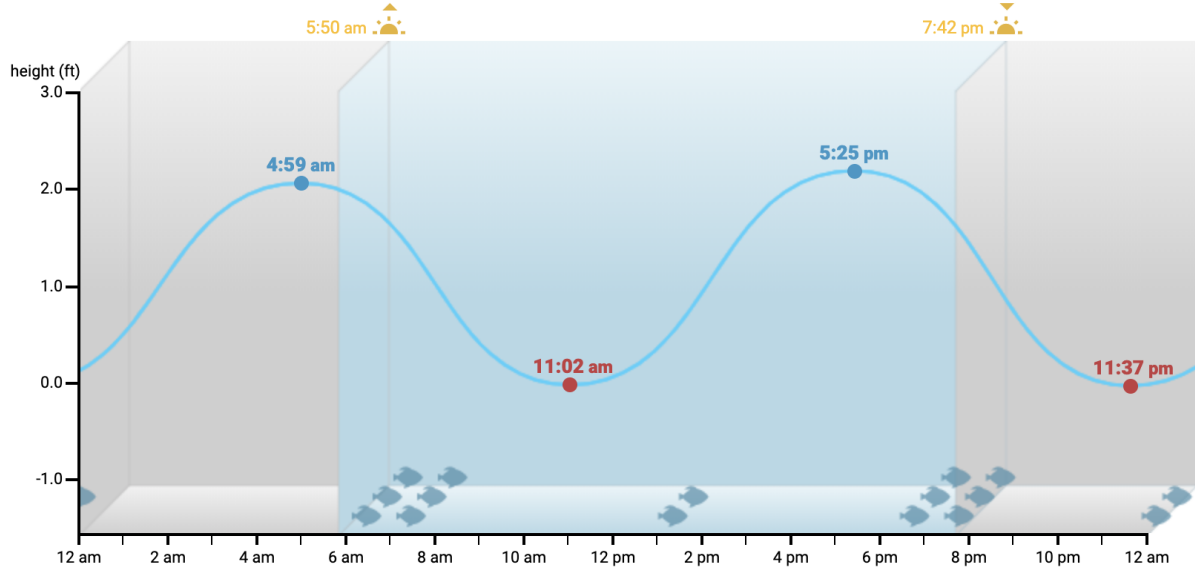


Figure C-1: Tidal time series for August 14, 2017. Times in EST (UTC -04:00). Low tide occurred at 11:02 EST / 15:02 UTC and high tide occurred at 17:25 EST / 21:25 UTC. The release occurred from 11:51 EST/ 15:51 UTC to 17:51 EST / 21:51 UTC. Data from <https://tides4fishing.com/> for the north side of Muskeget island.

through the drifter deployment, which occurred between 14:20 and 15:51. The evolution of the fields is shown in figure C-2 on page 199. The FTLE field at high tide (21:25 UTC) is also included for comparison.

### C.1.2 Spectral clustering

The parameter-free spectral clustering approach includes a sweep of radii  $r$  to determine the highest gap ratio. For the August 14, 2017 analyses, these sweeps in  $r$  are shown in figure C-3 on page 200, for the 14:20 to 20:20 and the 15:51 to 21:51 time windows. For each time window, the peaks are indicated on the figures. The results in chapter 4 correspond to these parameter settings: one peak was detected at 14:20 at  $r = 0.20$  and two peaks were detected at 15:51 at  $r = \{0.0125, 0.0425\}$ . For these values of  $r$ , the sweeps in offset coefficients are shown in figure C-4.

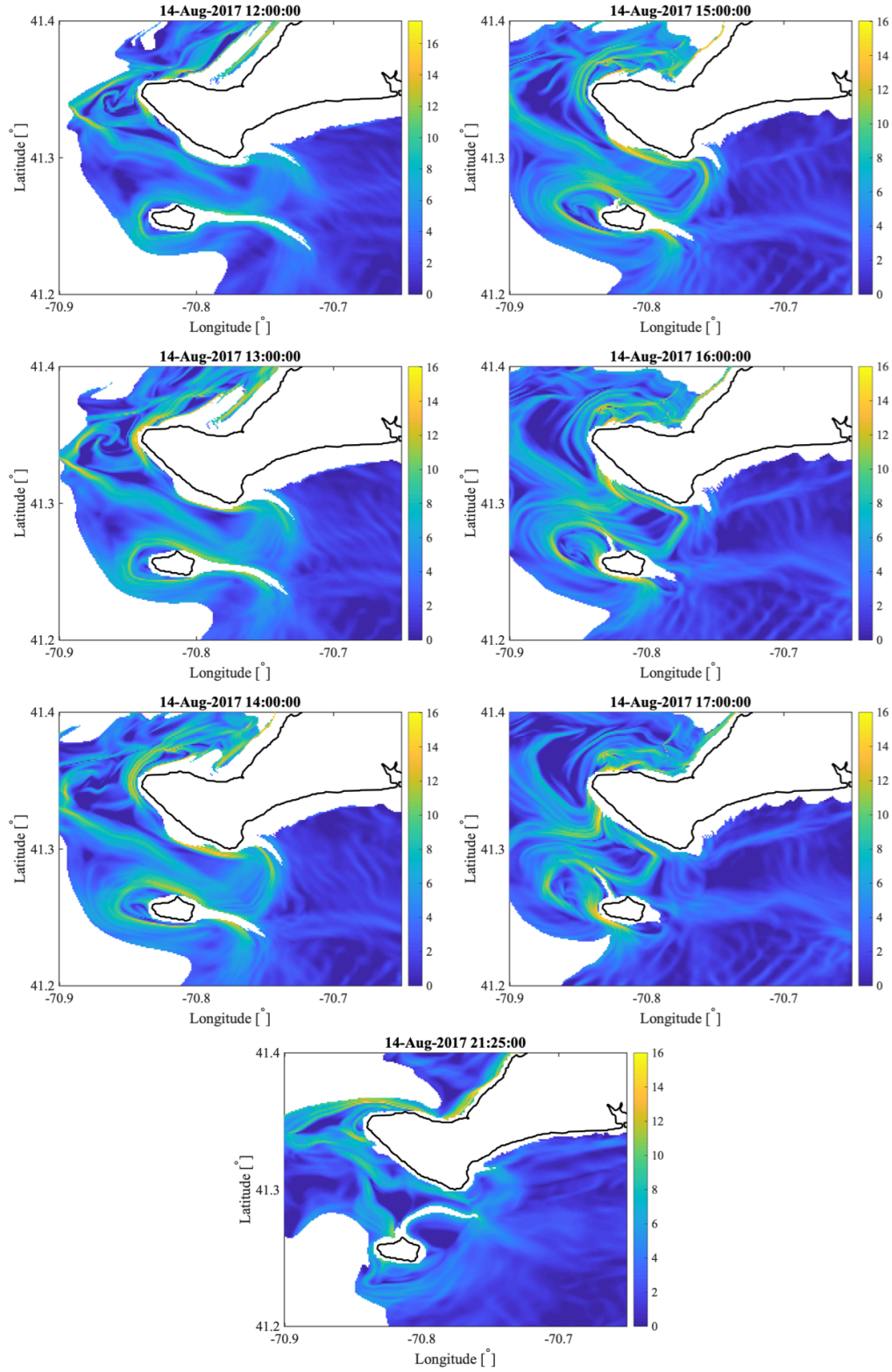
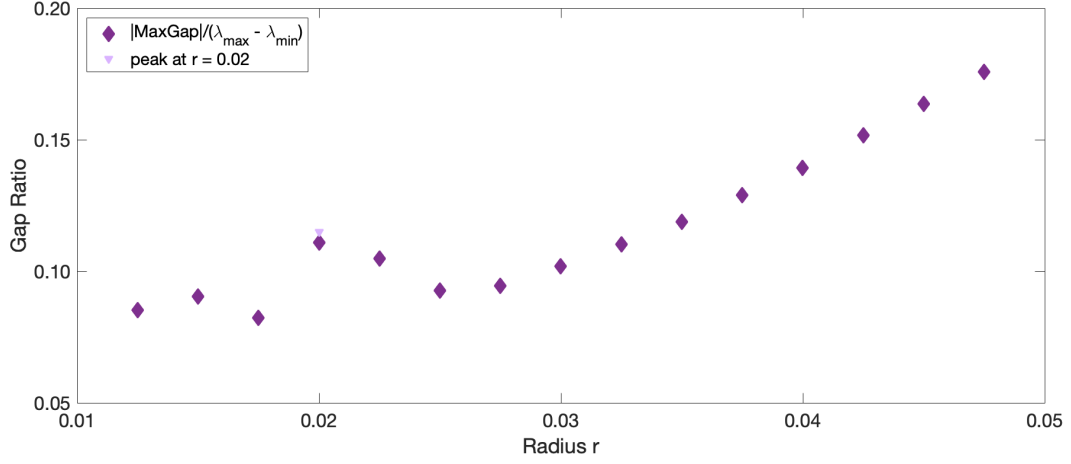
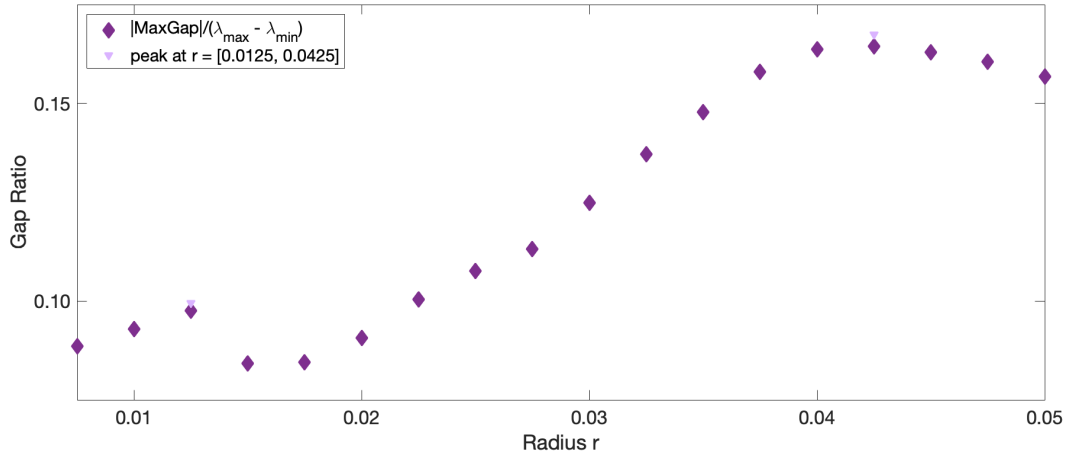


Figure C-2: Evolution of the FTLE field at different times of the tidal cycle for the day of August 14, 2017. The time window of integration was 6 hours for all runs. Low tide occurred at 15:02 and high tide occurred at 21:25. All times in UTC.

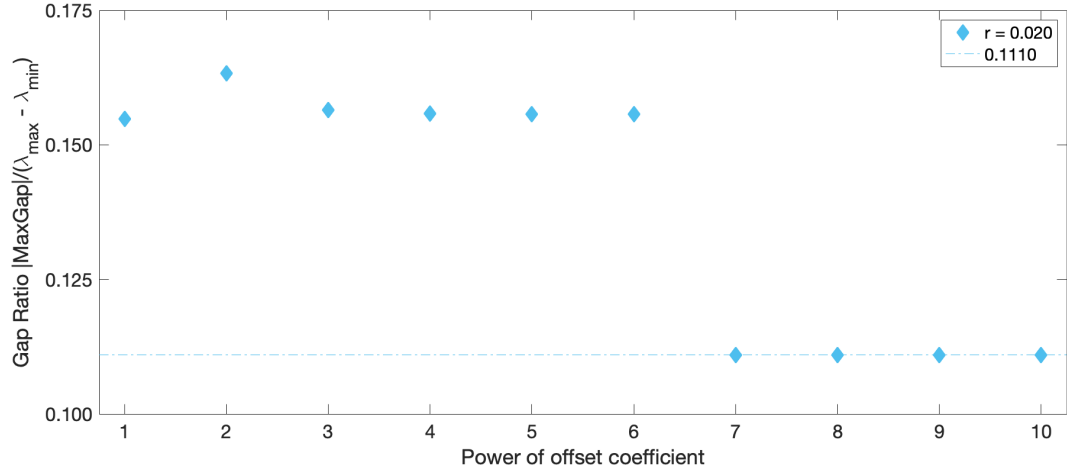


(a)

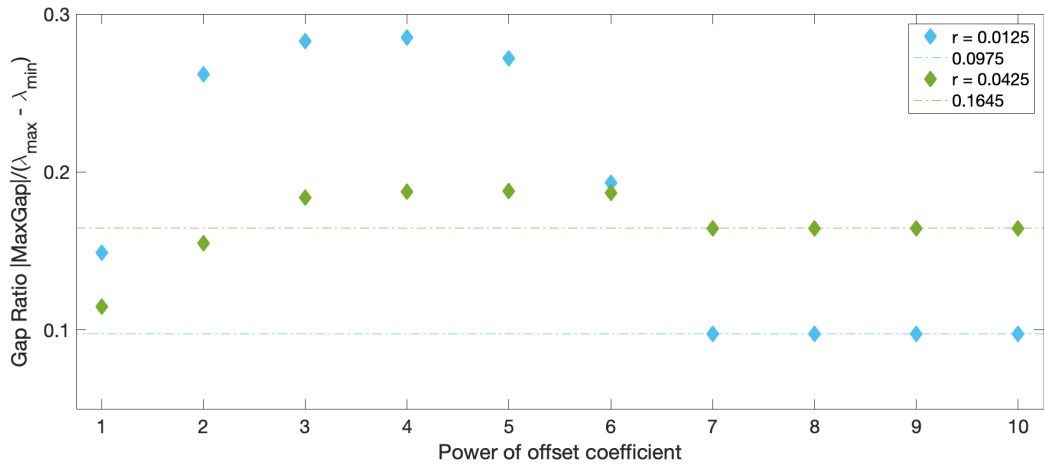


(b)

Figure C-3: Sweep of  $r$  radii for the parameter-free spectral clustering protocol used for the analysis of the 2017 No Man's Land experiment. (a) Sweep for the 14:20 to 20:20 time window. (b) Sweep for the 15:51 to 21:51 time window.



(a)



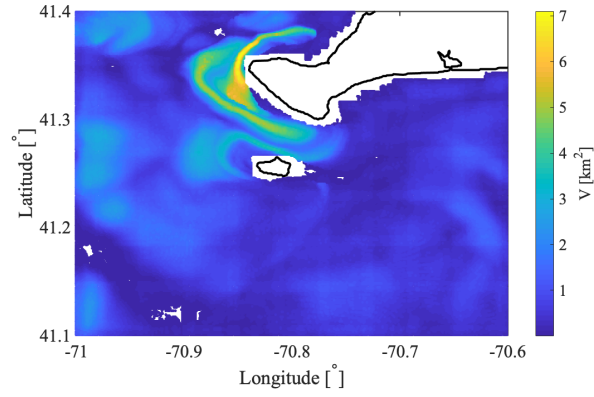
(b)

Figure C-4: Sweep of offset coefficients for the  $r$  values corresponding to the peaks in gap ratios from figure C-3. (a) Sweep for the 14:20 to 20:20 time window. (b) Sweep for the 15:51 to 21:51 time window.

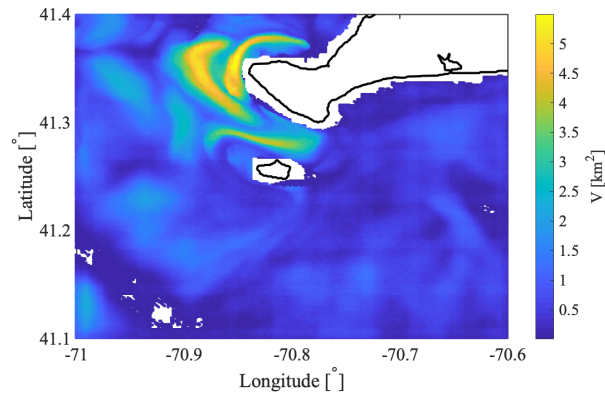
### C.1.3 Encounter volume

The encounter volume method is based on the number of trajectories that a reference trajectory encounters within a certain radius  $r$  over a chosen time window of integration. The  $r$  radius is chosen when the results converge below  $r$ : this number of trajectories encountered does not vary below a certain threshold. To obtain meaningful results, typically, the encounter volume must be computed over a larger domain. The full domains are shown in figure C-5 on page 203.

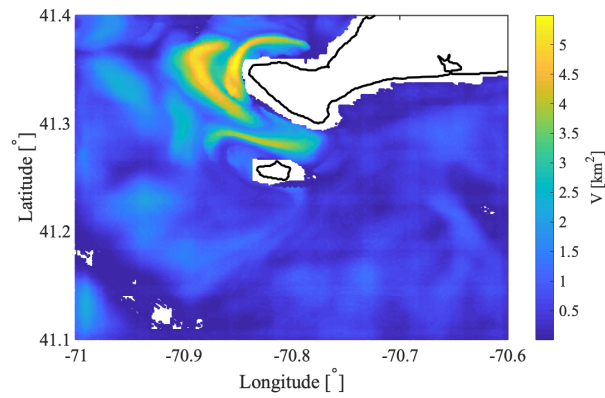




(a)



(b)



(c)

Figure C-5: Encounter volume over the entire domain of computations for the analysis of the 2017 No Man's Land experiment. (a) Results for the 14:20 to 20:20 time window. (b) Results for the 15:51 to 21:51 time window. (c) Results for the 21:51 to 03:51 time window.

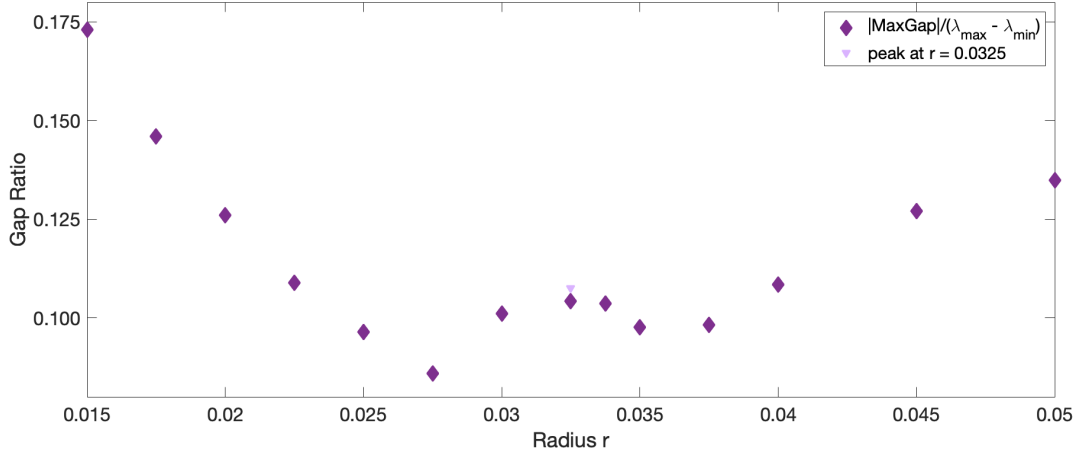


Figure C-6: Sweep of  $r$  radii for the parameter-free spectral clustering protocol used for the analysis of the 2018 No Man’s Land experiment. Sweep for the 04:00 to 10:00 time window.

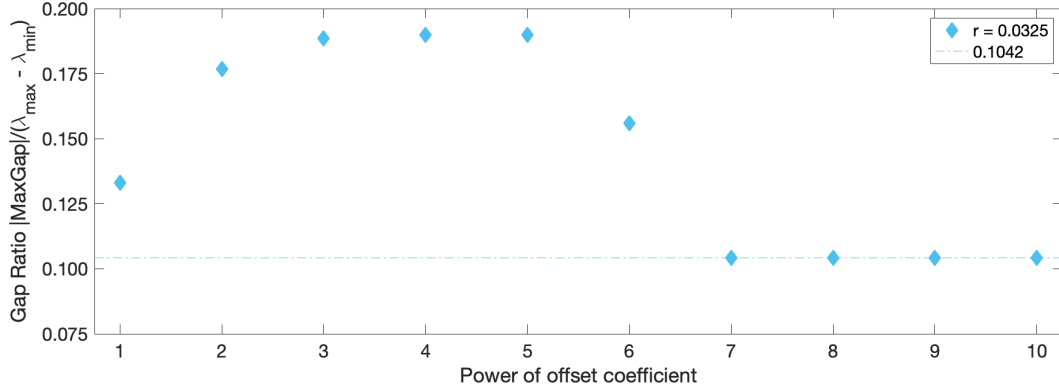
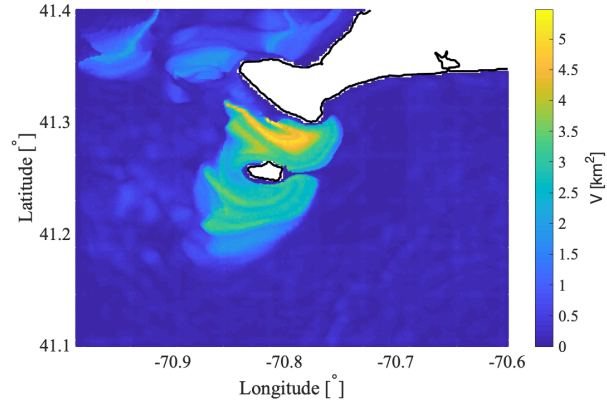


Figure C-7: Sweep of offset coefficients for the  $r$  value corresponding to the peak in gap ratios from figure C-6. Sweep for the 04:00 to 10:00 time window.

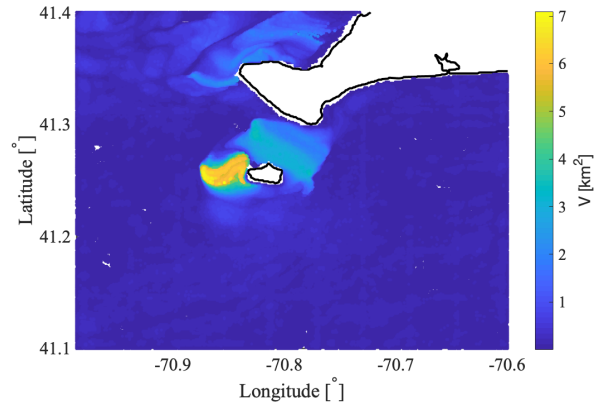
## C.2 2018 No Man’s Land Experiment

### C.2.1 Spectral clustering

The parameter-free spectral clustering protocol was followed for the August 8, 2018 analysis. The sweep in  $r$ -radii is shown in figure C-6 on page 204 for the 04:00 to 10:00 time window of integration. The peak in gap ratio was found at  $r = 0.0325$ . For this value of  $r$ , the sweep in offset coefficients is shown in figure C-7.



(a)



(b)

Figure C-8: Encounter volume over the entire domain of computations for the analysis of the 2018 No Man's Land experiment. (a) Results for the 04:00 to 10:00 time window. (b) Results for the 10:00 to 16:00 time window.

### C.2.2 Encounter volume

For the 04:00 to 10:00 time window on August 8, 2018, the full domains are shown in figure C-8 on page 205.



# Bibliography

- Ryan Abernathey & George Haller. 2018. Transport by Lagrangian vortices in the eastern pacific. *Journal of Physical Oceanography*, 48(3):667–685.
- Charu C. Aggarwal, Alexander Hinneburg & Daniel A. Keim. On the surprising behavior of distance metrics in high dimensional space. In Jan Van den Bussche & Victor Vianu, editors, *Database Theory — ICDT 2001*, pages 420–434, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg. ISBN 978-3-540-44503-6.
- Michael Allshouse & Jean-Luc Thiffeault. 2012. Detecting coherent structures using braids. *Physica D. Nonlinear Phenomena*, pages 95–105.
- Michael R. Allshouse & Thomas Peacock. August 2015a. Refining finite-time Lyapunov exponent ridges and the challenges of classifying them. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087410.
- Michael R. Allshouse & Thomas Peacock. 2015b. Lagrangian based methods for coherent structure detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(9):097617.
- Siavash Ameli & Shawn C. Shadden. 2019. A transport method for restoring incomplete ocean current measurements. *Journal of Geophysical Research: Oceans*, 124(1):227–242.
- Hassan Aref. 1984. Stirring by chaotic advection. *Journal of Fluid Mechanics*, 143:1–21.
- V. Artale, G. Boffetta, A. Celani, M. Cencini & A. Vulpiani. 1997. Dispersion of passive tracers in closed basins: Beyond the diffusion coefficient. *Physics of Fluids*, 9:3162–3171.

- E. Aurell, G. Boffetta, G. Crisanti, A. Paladin & A. Vulpiani. 1997. Predictability in the large: An extension of the concept of Lyapunov exponent. *J. Phys. A: Math. Gen.*, 30: 1–26.
- Daniel Beltrá. A ship drifts amidst a heavy band of oil spilled in the gulf of mexico from the deepwater horizon wellhead, 2010. URL <https://danielbeltra.photoshelter.com/portfolio/G0000N9uDgKewQWk/I00002t9n.7rAla0>.
- Francisco J. Beron-Vera, Maria Josefina Olascoaga, George Haller, Mohammad Farazmand, Joaquin Triñanes & Yan Wang. 2015. Dissipative inertial transport patterns near coherent Lagrangian eddies in the ocean. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087412.
- Francisco J. Beron-Vera, Alireza Hadjighasem, Qiong Xia, Maria Josefina Olascoaga & George Haller. 2018. Coherent Lagrangian swirls among submesoscale motions. *Proceedings of the National Academy of Sciences, Proceedings of the National Academy of Sciences*.
- L. W. Botsford, J. W. White, M.-A. Coffroth, Claire B. Paris, S. Planes, T. L. Shearer, S. R. Thorrold & G. P. Jones. Jun 2009. Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs*, 28(2):327–337.
- Zeynel Cebeci & Figen Yildiz. 2015. Comparison of k-means and fuzzy c-means algorithms on different cluster structures. *Journal of Agricultural Informatics*, 6(3):13–23.
- Timothy J. Crone & Maya Tolstoy. 2010. Magnitude of the 2010 gulf of mexico oil leak. *Science*, 330(6004):634–634.
- Russ E. Davis. 1983. Oceanic property transport, Lagrangian particle statistics, and their prediction. *Journal of Marine Research*, 41:163–194.
- Russ E. Davis. 1985. Drifter observations of coastal surface currents during code: The method and descriptive view. *Journal of Geophysical Research: Oceans*, 90(C3):4741–4755.

- Eric Deleersnijder & Pierre F. J. Lermusiaux. 2008. Multi-scale modeling: nested-grid and unstructured-mesh approaches. *Ocean Dynamics*, 58:335–336.
- Pedro Domingos. October 2012. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87.
- Mohammad Farazmand & George Haller. 2012. Computing Lagrangian Coherent Structures from their variational theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(1):013128.
- Mohammad Farazmand & George Haller. 2016. Polar rotation angle identifies elliptic islands in unsteady dynamical systems. *Physica D*, 315:1–12.
- Mohammad Farazmand, Daniel Blazevski & George Haller. 2014. Shearless transport barriers in unsteady two-dimensional flows and maps. *Physica D: Nonlinear Phenomena*, 278-279:44 – 57.
- Mohammad Farazmand, George Haller & Florian Huhn. 2016. Defining coherent vortices objectively from the vorticity. *Journal of Fluid Mechanics*, 795:136–173.
- Raffaele Ferrari & Carl Wunsch. 2009. Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, 41(1):253–282.
- Christopher B. Field, Michael J. Behrenfeld, James T. Randerson & Paul Falkowski. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, 281(5374):237–240.
- Lermusiaux Pierre F.J., Jens Schröter, Sergey Danilov, Mohamed Iskandarani, Nadia Pinardi & Joannes J. Westerink. 2013. Multiscale modeling of coastal, shelf and global ocean dynamics. *Ocean Dynamics*, 63:1341–1344.
- Taryn Foster & James Gilmour. jul 2018. Reproduction of brooding corals at scott reef, western australia. *Matters, Matters*.
- Gary Froyland. 2013. An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. *Physica D: Nonlinear Phenomena*, 250:1 – 19.

- Gary Froyland & Kathrin Padberg-Gehle. 2015. A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087406.
- Soumi Ghosh & Sanjay Kumar Dubey. 2013. Comparative analysis of k-means and fuzzy cmeans algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4):35–39.
- James P. Gilmour, Luke D. Smith, Andrew J. Heyward, Andrew H. Baird & Morgan S. Pratchett. 2013. Recovery of an isolated coral reef system following severe disturbance. *Science*, 340(6128):69–71.
- Rebecca H. Green, Nicole L. Jones, Matthew D. Rayson, Ryan J. Lowe, Cynthia E. Bluteau & Gregory N. Ivey. 2019. Nutrient fluxes into an isolated coral reef atoll by tidally driven internal bores. *Limnology and Oceanography*, 64(2):461–473.
- Alireza Hadjighasem & George Haller. 2016. Geodesic transport barriers in jupiter’s atmosphere: A video-based analysis. *SIAM Review*, 58(1):69–89.
- Alireza Hadjighasem, Daniel Karrasch, H. Teramoto & George Haller. 2016. Spectral-clustering approach to Lagrangian vortex detection. *Physical Review E*, 93:063107.
- Alireza Hadjighasem, Mohammad Farazmand, Gary Froyland & George Haller. 2017. A critical comparison of Lagrangian methods for coherent structure detection. *CHAOS*, 27: 053104.
- Patrick J. Haley & Pierre F. J. Lermusiaux. Dec 2010. Multiscale two-way embedding schemes for free-surface primitive equations in the “multidisciplinary simulation, estimation and assimilation system”. *Ocean Dynamics*, 60(6):1497–1537.
- Patrick J. Haley, Arpit Agarwal & Pierre F.J. Lermusiaux. 2015. Optimizing velocities and transports for complex coastal regions and archipelagos. *Ocean Modelling*, 89:1 – 28.
- G. Haller & A.C. Poje. 1998. Finite time transport in aperiodic flows. *Physica D: Nonlinear Phenomena*, 119(3):352 – 380.



- George Haller. 2015. Lagrangian Coherent Structures. *Annual Review of Fluid Mechanics*, 47(1):137–162.
- George Haller & Francisco J. Beron-Vera. 2013. Coherent Lagrangian vortices: The black holes of turbulence. *Journal of Fluid Mechanics*, 731(R4).
- George Haller & G. Yuan. 2000. Lagrangian Coherent Structures and mixing in two-dimensional turbulence. *Physica D: Nonlinear Phenomena*, 147(3):352 – 370.
- Angelique C. Haza, A. Griffa, P. Martin, A. Molcard, Tamay M. Özgökmen, Andrew C. Poje, R. Barbanti, J.W. Book, P.-M. Poulain, M. Rixen & P. Zanasca. 2007. Tmodel-based directed drifter launches in the adriatic sea: Results from the dart experiment. *Geophysical Research Letters*, 34:L10605.
- Angelique C. Haza, Tamay M. Özgökmen, Annalisa Griffa, A. Molcard, P.-M. Poulain & G. Peggion. 2010. Transport properties in small-scale coastal flows: relative dispersion from vhf radar measurements in the gulf of la spezia. *Ocean Dynamics*, 60:861–882.
- T. P. Hughes, M. L. Barnes, D. R. Bellwood, J. E. Cinner, G. S. Cumming, J. B. C. Jackson, J. Kleypas, I. A. van de Leemput, J. M. Lough, T. H. Morrison, S. R. Palumbi, E. H. van Nes & M. Scheffer. 2017. Coral reefs in the anthropocene. *Nature*, 546:82–90.
- Gregg A. Jacobs, Brent P. Bartels, Darek J. Bogucki, Francisco J. Beron-Vera, Shuyi S. Chen, Emanuel F. Coelho, Milan Curcic, Annalisa Griffa, Matthew Gough, Brian K. Haus, Angelique C. Haza, Robert W. Helber, Patrick J. Hogan, Helga S. Huntley, Mohamed Iskandarani, Falko Judt, Albert D. Kirwan, Nathan Laxague, Arnolfo Valle-Levinson, Bruce L. Lipphardt, Arthur J. Mariano, Hans E. Ngodock, Guillaume Novelli, María J. Olascoaga, Tamay M. Özgökmen, Andrew C. Poje, Ad J.H.M. Reniers, Clark D. Rowley, Edward H. Ryan, Scott R. Smith, Peter L. Spence, Prasad G. Thoppil & Mozheng Wei. 2014. Data assimilation considerations for improved ocean predictability during the Gulf of Mexico Grand Lagrangian Deployment (GLAD). *Ocean Modelling*, 83(Supplement C): 98 – 117.

- S. E. Kahng, J. R. Garcia-Sais, H. L. Spalding, E. Brokovich, D. Wagner, E. Weil, L. Hinderstein & R. J. Toonen. 2010. Community ecology of mesophotic coral reef ecosystems. *Coral Reefs*, 29:255–275.
- Sam Kahng, Joshua M. Copus, editor="Rossi-Sergio Wagner, Daniel", Lorenzo Bramanti, Andrea Gori & Covadonga Orejas. 2017. *Mesophotic Coral Ecosystems*, pages 185–206. Springer International Publishing, Cham, 2017. ISBN 978-3-319-21012-4.
- Anthony Kirincich. 2016. Remote sensing of the surface wind field over the coastal ocean via direct calibration of HF radar backscatter power. *Journal of Atmospheric and Oceanic Technology*, 33(7):1377–1392.
- S.P. Kolinski & E.F. Cox. 2003. An update on modes and timing of gamete and planula release in Hawaiian scleractinian corals with implications for conservation and management. *Pacific Science*, 57:17–27.
- Johnathan T. Kool, Claire B. Paris, Serge Andréfouët & Robert K. Cowen. 2010. Complex migration and the development of genetic structure in subdivided populations: an example from caribbean coral reef ecosystems. *Ecography*, 33:597–606.
- Johnathan T. Kool, Claire B. Paris, Paul H. Barber & Robert K. Cowen. 2011. Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography*, 20(5).
- Andrew S. Kough, Claire B Paris-Limouzy & Mark J. Butler. 1 2018. The spatial context of “winning” in MPA network design: Location matters. *Conservation Letters, Conservation Letters*.
- Joseph H. LaCasce. 2008. Statistics from Lagrangian observations. *Progress in Oceanography*, 77(1):1 – 29.
- Yoav Lehahn, Francesco d’Ovidio & Ilan Koren. 2018. A satellite-based Lagrangian view on phytoplankton dynamics. *Annual Review of Marine Science*, 10(1):99–119. PMID: 28961072.

- Ryan J. Lowe & James L. Falter. sept 2015. Oceanic forcing of coral reefs. *Annual Review of Marine Science*, 7:43–66.
- Amala Mahadevan. 2016. The impact of submesoscale physics on primary productivity of plankton. *Annu. Rev. Mar. Sci.*, 8:161–84.
- Ana M. Mancho, Stephen Wiggins, Jezabel Curbelo & Carolina Mendoza. 2013. Lagrangian descriptors: A method for revealing phase space structures of general time dependent dynamical systems. *Communications in Nonlinear Science and Numerical Simulation*, 18: 3530–3557.
- Irina Marinov, Michael Follows, Anand Gnanadesikan, Jorge L. Sarmiento & Richard D. Slater. 2008. How does ocean biology affect atmospheric pCO<sub>2</sub>? theory and models. *Journal of Geophysical Research: Oceans*, 113(C7).
- Mani Mathur, George Haller, Thomas Peacock, Jori Elan Ruppert-Felsot & Harry L. Swinney. April 2007. Uncovering the Lagrangian Skeleton of Turbulence. *Physical Review Letters*, 98(14):144502.
- James C. McWilliams. 1976. Maps from the mid-ocean dynamics experiment: Part i. geostrophic streamfunction. *Journal of Physical Oceanography*, 6(6):810–827.
- James C. McWilliams. 2016. Submesoscale currents in the ocean. *Proceedings of the Royal Society A*, 472:1–17.
- Carolina Mendoza & Anna M. Mancho. 2012. Review article: “the Lagrangian description of aperiodic flows: a case study of the Kuroshio current”. *Nonlinear Processes in Geophysics*, 19(4):449–472.
- Stephen G. Monismith. 2007. Hydrodynamics of coral reefs. *Annual Review of Fluid Mechanics*, 39:37–55.
- María J. Olascoaga & George Haller. 2012. Forecasting sudden changes in environmental pollution patterns. *Proceedings of the National Academy of Sciences*, 109(13):4738–4743.

- María J. Olascoaga, Francisco J. Beron-Vera, George Haller, Joaquan Triñanes, M. Iskandarani, Emanuel F. Coelho, Brian K. Haus, Helga S. Huntley, Gregg A. Jacobs, A. D. Kirwan, Bruce L. Lipphardt, T. M. Özgökmen, A. J. H. M. Reniers & A. Valle-Levinson. 2013. Drifter motion in the gulf of mexico constrained by altimetric Lagrangian coherent structures. *Geophysical Research Letters*, 40(23):6171–6175.
- Kristjan Onu, Florian Huhn & George Haller. 2015. LCS tool: A computational platform for Lagrangian Coherent Structures. *Journal of Computational Science*, 7:26 – 36.
- Thomas Peacock & George Haller. January 2013. Lagrangian Coherent Structures: The hidden skeleton of fluid flows. *Physics Today*, 66(2):41–47.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot & E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Andrew C. Poje, Tamay M. Özgökmen, Bruce L. Lipphardt, Brian K. Haus, Edward H. Ryan, Angelique C. Haza, Gregg A. Jacobs, A. J. H. M. Reniers, María J. Olascoaga, Guillaume Novelli, Annalisa Griffa, Francisco J. Beron-Vera, Shuyi S. Chen, Emanuel F. Coelho, Patrick J. Hogan, Albert D. Kirwan, Helga S. Huntley & Arthur J. Mariano. 2014. Submesoscale dispersion in the vicinity of the Deepwater Horizon spill, 111(35): 12693–12698.
- James Poupard & Benjamin Franklin. A chart of the gulf stream, 1786. URL <https://lccn.loc.gov/2004627238>.
- Matthew D. Rayson. *The tidally-driven ocean dynamics of the Browse Basin and Kimberley Shelf region, Western Australia*. PhD thesis, The University of Western Australia, School of Environmental Systems Engineering, December 2011.
- Matthew D. Rayson, Gregory N. Ivey, Nicole L. Jones, Michael J. Meuleners & Geoffrey W. Wake. Internal tide dynamics in a topographically complex region: Browse basin, australian north west shelf. *Journal of Geophysical Research: Oceans*, 116(C1).

- Matthew D. Rayson, Gregory N. Ivey, Nicole L. Jones & Oliver B. Fringer. 2018. Resolving high-frequency internal waves generated at an isolated coral atoll using an unstructured grid ocean model. *Ocean Modelling*, 122:67 – 84.
- Hanny E. Rivera. *Genetic connectivity, adaptation, and phenotypic plasticity of corals and anemones under thermal stress*. PhD thesis, Massachusetts Institute of Technology/Woods Hole Oceanographic Institution, 2019.
- Irina I. Rypina. *Lagrangian Coherent Structures and Transport in Two-Dimensional Incompressible Flows with Oceanographic and Atmospheric Applications*. PhD thesis, University of Miami, 2007.
- Irina I. Rypina & Lawrence J. Pratt. 2017. Itrajectory encounter volume as a diagnostic of mixing potential in fluid flows. *Nonlinear Processes Geophysics*, 24:189–202.
- Irina I. Rypina, Michael G. Brown, Francisco. J. Beron-Vera, Hüseyin Koçak, María J. Olascoaga & Ilya A. Udovydchenkov. 2007. On the Lagrangian dynamics of atmospheric zonal jets and the permeability of the stratospheric polar vortex. *Journal of the Atmospheric Sciences*, 64(10):3595–3610.
- Irina I. Rypina, S. E. Scott, Lawrence J. Pratt & Michael G. Brown. 2011. Investigating the connection between complexity of isolated trajectories and Lagrangian Coherent Structures. *Nonlinear Processes in Geophysics*, 18(6):977–987.
- Irina I. Rypina, Steven R. Jayne, Sachiko Yoshida, Alison M. Macdonald, Elizabeth Douglass & Ken Buesseler. 2013. Short-term dispersal of Fukushima-derived radionuclides off japan: modeling efforts and model-data intercomparison. *Biogeosciences*, 10(7):4973–4990.
- Irina I. Rypina, S. R. Jayne, S. Yoshida, A. M. Macdonald & K. Buesseler. 2014a. Drifter-based estimate of the 5 year dispersal of fukushima-derived radionuclides. *Journal of Geophysical Research: Oceans*, 119(11):8177–8193.
- Irina I. Rypina, Anthony R. Kirincich, R. Limeburner & Ilya A. Udovydchenkov. 2014b. Eulerian and Lagrangian correspondence of high-frequency radar and surface drifter data:

- Effects of radar resolution and flow components. *Journal of Atmospheric and Oceanic Technology*, 31(4):945–966.
- Irina I. Rypina, Anthony Kirincich, Steven Lentz & Miles A. Sundermeyer. 2016. Investigating the eddy diffusivity concept in the coastal ocean. *Journal of Physical Oceanography*, 46(7):2201–2218.
- Irina I. Rypina, Stefan G. Llewellyn Smith & Lawrence J. Pratt. 2018. Connection between encounter volume and diffusivity in geophysical flows. *Nonlinear Processes in Geophysics*, 25(2):267–278.
- Mathias Sandulescu, Cristóbal López, Emilio Hernández-García & Ulrike Feudel. 2007. Plankton blooms in vortices: The role of biological and hydrodynamic time scales. *Nonlinear Processes in Geophysics*, 14:443–454.
- Carmine Senatore & Shane D. Ross. Detection and characterization of transport barriers in complex flows via ridge extraction of the finite time Lyapunov exponent field. *International Journal for Numerical Methods in Engineering*, 86(9):1163–1174.
- Seongbong Seo, Young-Gyu Park, Chanhung Jeon, Hong Sik Min, Dong Guk Kim & Jae-Hun Park. 2018. Turbulent mixing induced by cold water intrusion over the southwestern east sea. *Journal of Coastal Research*, 85(SI):476–480.
- Mattia Serra & George Haller. 2016. Objective Eulerian Coherent Structures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(5):053110.
- Shawn C. Shadden, Francois Lekien & Jerrold E. Marsden. 2005. Definition and properties of Lagrangian Coherent Structures from Finite-Time Lyapunov Exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena*, 212(3):271 – 304.
- Huan-Jie Shao, Ruo-Shan Tseng, Ren-Chieh Lien, Yu-Chia Chang & Jau-Ming Chen. 2018. Turbulent mixing on sloping bottom of an energetic tidal channel. *Continental Shelf Research*, 166:44 – 53.

- Jianbo Shi & J. Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.
- Maïke Sonnewald, Carl Wunsch & Patrick Heimbach. 2019. Unsupervised learning reveals geography of global ocean dynamical regions. *Earth and Space Science*, 0(ja).
- Hui Sun, Qingxuan Yang & Jiwei Tian. Oct 2018. Microstructure measurements and finescale parameterization assessment of turbulent mixing in the northern south china sea. *Journal of Oceanography*, 74(5):485–498.
- Jean-Luc Thiffeault. 2010. Braids of entangled particle trajectories. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(1):017516–017514.
- Leif N. Thomas, Amit Tandon & Amala Mahadevan. 2008. *Submesoscale Processes and Dynamics*, pages 17–38. American Geophysical Union (AGU), 2008. ISBN 9781118666432.
- Erik van Sebille, Stephen M. Griffies, Ryan Abernathey, Thomas P. Adams, Pavel Berloff, Arne Biastoch, Bruno Blanke, Eric P. Chassignet, Yu Cheng, Colin J. Cotter, Eric Deleersnijder, Kristofer Ds, Henri F. Drake, Sybren Drijfhout, Stefan F. Gary, Arnold W. Heemink, Joakim Kjellsson, Inga Monika Koszalka, Michael Lange, Camille Lique, Graeme A. MacGilchrist, Robert Marsh, C. Gabriela Mayorga Adame, Ronan McAdam, Francesco Nencioli, Claire B. Paris, Matthew D. Piggott, Jeff A. Polton, Siren Rhs, Syed H.A.M. Shah, Matthew D. Thomas, Jinbo Wang, Phillip J. Wolfram, Laure Zanna & Jan D. Zika. 2018. Lagrangian ocean analysis: Fundamentals and practices. *Ocean Modelling*, 121:49 – 75.
- Guilherme Salvador Vieira & Michael R. Allshouse. Uncertainty quantification in Lagrangian clustering analysis. In *APS Meeting Abstracts*, page F48.00006, 2019.
- Emmanuel Villerraux. 2019. Mixing versus stirring. *Annual Review of Fluid Mechanics*, 51(1):245–273.
- Ulrike von Luxburg. Dec 2007. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

- Stephen Wiggins. 2005. The dynamical system approach to Lagrangian transport in oceanic flows. *Annual Review of Fluid Mechanics*, 37:295–328.
- Matthew O. Williams, Irina I. Rypina & Clarence W. Rowley. 2015. Identifying finite-time coherent sets from limited quantities of Lagrangian data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087408.
- Sally Wood, Claire B. Paris, Andy Ridgwell & Erica J. Hendy. 2014. Modelling dispersal and connectivity of broadcast spawning corals at the global scale. *Global Ecology and Biogeography*, 23(1):1–11.
- Carl Wunsch & Raffaele Ferrari. 2018. 100 years of the ocean general circulation. *Meteorological Monographs*, 59:7.1–7.32.
- Jiangtao Xu, Ryan J. Lowe, Gregory N. Ivey, Nicole L. Jones & Zhenling Zhang. 2018. Contrasting heat budget dynamics during two La niña marine heat wave events along Northwestern Australia. *Journal of Geophysical Research C: Oceans*, 123(2):1563–1581.
- Zhengguang Zhang & Bo Qiu. Evolution of submesoscale ageostrophic motions through the life cycle of oceanic mesoscale eddies. *Geophysical Research Letters*, 45(21):11,847–11,855.
- Zhengguang Zhang, Wei Wang & Bo Qiu. 2014. Oceanic mass transport by mesoscale eddies. *Science*, 345:322–324.